Micropitting and Related Phenomena in Case Carburised Gears

Adrian Oila

A Thesis Submitted for the Degree of
Doctor of Philosophy

July 2003

DECLARATION

This thesis records work carried out in the School of Chemical Engineering and Advanced Materials and the Design Unit – Gear Technology Centre at the University of Newcastle upon Tyne between October 1999 and November 2002 and is original except where acknowledged by reference to work of others.

Several sections of this work have already been published (or are in press) in different journals, and these papers are summarised below.

1. Oila, A. & Bull, S.J. (2003) Nanoindentation testing of gear steels. *Zeitschrift für Metallkunde* 94 (7): 793-797.

No portion of this work is being, nor has been, submitted for a degree, diploma or any other qualification at any other university.

ACKNOWLEDGEMENTS

I would like to express my appreciation for all the efforts of everyone who contributed in some way to this project.

First and foremost, I would like to thank my supervisor, Professor Steve Bull for his patience, understanding and moral support and for supporting this work with ideas and criticism. His guidance throughout my research and write-up was invaluable. His constant encouragement gave me a boost of confidence and stimulation. Without his support this PhD would not have been possible.

I wish to thank Professor Trevor Page for his advices and support in various occasions.

I also owe a debt of gratitude to Dr. Brian Shaw for many stimulating discussions and helpful comments on this research. Thanks also to the members of Materials Laboratory from the Design Unit, Chris Aylott and Phil Wilson for the cooperative and friendly environment they created which supported me throughout the last stage of this research.

This thesis would have not been possible without the help of the technicians from the Materials Division. My grateful thanks go to Malcolm Black, Ken Madden, Steve Charlton, Collin Henry and Alf Nessbit.

I wish to thank my parents in Romania who have been very supportive throughout all the stages of my work, and for understanding my absence in some moments when they needed me. This thesis is dedicated to them.

The final acknowledgement goes to those who funded this work. My studies were supported by the Newcastle University Research Committee and Caterpillar Inc. Corporation. Special thanks to the Design Unit – Gear Technology Centre for the financial support during the last stage of this project.

ABSTRACT

Micropitting is a form of surface contact fatigue encountered in bearings and gears, under lubricating conditions, which lead to their premature failure. All gears are susceptible to micropitting, including spur, helical and bevel. Micropitting can occur with all heat treatments applied to gears and with both, synthetic and mineral lubricants. It can occur after a relatively short period of operation and, after a certain number of cycles, gears need to be replaced due to the increased noise and vibrations caused by the deviations of the tooth profile. Continuing operation of affected gears can lead to a catastrophic type of failure (i.e., tooth breakage). These considerations explain the increasing current interest in micropitting.

It has been reported that micropitting in bearings is associated with a specific microstructural transformation in steel, i.e. martensite decay. However, to the author's knowledge, this transformation has not been reported in gears. In the present work, extensive metallurgical investigations have been carried out and they revealed that the same transformation occurs in gears.

The aim of this project was to describe the mechanism of micropitting by taking into account the influence of several controlling factors such as, material, surface finish, lubricant, load, temperature, speed and, slide-to-roll ratio. Their influence is assessed with a fractional factorial experimental design. Several non-destructive techniques have been used in order to monitor the specimen condition during and after running, such as X-ray diffraction, optical profilometry, light microscopy. The mechanical properties of the products of martensite decay, known as dark etching regions, white etching bands and butterflies are highly relevant to the fatigue behaviour of the steel. Nanoindentation and AFM techniques have been used to determine these properties.

A micropitting mechanism correlated with the mechanism of martensite decay in gears is suggested based on these analyses.

NOMENCLATURE

A_c (m^2)	area of contact,
a, b (m)	semi-axes of the contact ellipse,
d (μm/cycle)	DER transformation factor,
E (N/mm^2)	Young's modulus,
E' (N/mm^2)	elastic contact modulus,
G	dimensionless materials parameter,
g (m/s^2)	acceleration due to gravity,
H (GPa)	hardness,
h_{cen}	dimensionless central film thickness,
h_{min}	dimensionless minimum film thickness,
M (%)	micropitting parameter,
M_s (°C)	martensite start temperature,
m (kg)	mass,
N_{px}	number of peaks in Ox direction,
N_{py}	number of peaks in Oy direction,
N_s	number of summits,
P (N)	applied load,
p (μm/cycle)	PDL transformation factor,
p_0 (N/m^2)	maximum contact pressure,
R (m)	radius of curvature of the equivalent cylinder,
R', R'' (m)	major and minor relative radii of curvature,
R_a (μm)	arithmetic average roughness,
R_e (m)	effective radius of curvature,
R_p (μm)	maximum peak height,
R_q (μm)	square root roughness,
R_t (μm)	peak to valley distance,
R_v (μm)	maximum valley depth,

R^*_q	composite roughness,
r_p (μm)	mean peak height,
r_s (μm)	mean summit height,
S	slide-to-roll ratio,
U	dimensionless speed parameter,
u (m/s)	linear velocity,
W_L	dimensionless load parameter,
w (μm/cycle)	WEB transformation factor,
α (Pa^{-1})	pressure-viscosity coefficient,
β^* (μm)	mean summit curvature,
δ (m)	the approach of distant points,
δ_c (mm)	contact depth,
Δ_a (mrad)	arithmetical mean slope of the roughness profile,
Δ_q or σ_m (mrad)	root-mean-square slope of the roughness profile,
η (Pa·s)	dynamic viscosity,
Λ	wavelength ratio,
λ	specific film thickness (lambda ratio),
λ_a	arithmetic average wavelength of the roughness profile,
λ_q	root mean square wavelength of the roughness profile,
μ	friction coefficient,
μ_s	summit density,
ν	Poisson's ratio,
ρ (μm)	mean radius of curvature of the roughness profile,
σ_k (μm)	root-mean-square radius of curvature of the profile,
σ_p (μm)	standard deviation of the peak height,
σ_s (μm)	standard deviation of the summit height,
τ (Pa)	shear stress.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

Chapter 9: DISCUSSION AND CONCLUSIONS

~ Chapter 1 ~

INTRODUCTION

1 INTRODUCTION

1.1 What is micropitting?

Gears have been used since the dawn of engineering. Early gears were made from wood and were often lubricated with animal fat grease. The industrial revolution in the eighteenth century saw an explosion in the use of metal gearing. A science of gear design and manufacture rapidly developed through the nineteenth century. Modern metallurgy has greatly increased the life of industrial and automotive gears. Nowadays gears are manufactured from a large variety of materials but steels are the most widely used due to their good mechanical properties and relatively low price. During the post-world war II era case carburised steels have been increasingly used for high performance gear applications due to a beneficial combination of properties (i.e., surface hardness and core strength).

Failure of the gear tooth can occur in service when bending stresses or contact stresses exceed the safety limits. Gear failures can cost thousands of pounds in lost production and replacement costs. For high-speed, high-load machinery such as in marine transmission, premature failure of gears represents an important issue. Two or more failure modes may occur simultaneously, or one may be the result of the other. Failure due to surface contact fatigue represents an important fraction of all gear failure types (about 20% according to ASM, 1975). Moreover, other types of gear failure (e.g., tooth breakage) can have their origin in a form of surface contact fatigue.

All the surface contact fatigue phenomena (i.e., *pitting*, *flaking*, *spalling*, *case crushing* and, *micropitting*) are characterised by the removal of metal and the formation of cavities on the gear tooth surface due to the initiation and propagation of cracks.
A fatigue crack can nucleate at or near the surface. Sub-surface cracks usually initiate at inhomogeneties such as non-metallic inclusions or carbides present in the steel

microstructure. By improving the cleanliness of the steel, steel makers have eliminated the larger non-metallic inclusions (i.e., oxides) that previously provided easy nucleation sites for sub-surface cracks. However, the introduction of clean steels has the effect of bringing surface nucleated fatigue to the forefront.

Regions with grey, dull appearance as if they were etched have been observed on gear tooth flanks, particularly on dedendum for a long time. Known as the "grey staining" this phenomenon was regarded in the past as a secondary wear problem or even as beneficial because of the surface finish improvement. When the grey staining regions are observed under higher magnification (see Figure 1-1) they resemble macroscopic pitting but on a smaller scale, hence the term *micropitting*. The phenomenon is also referred to as frosting, glazing, microspalling, or peeling.

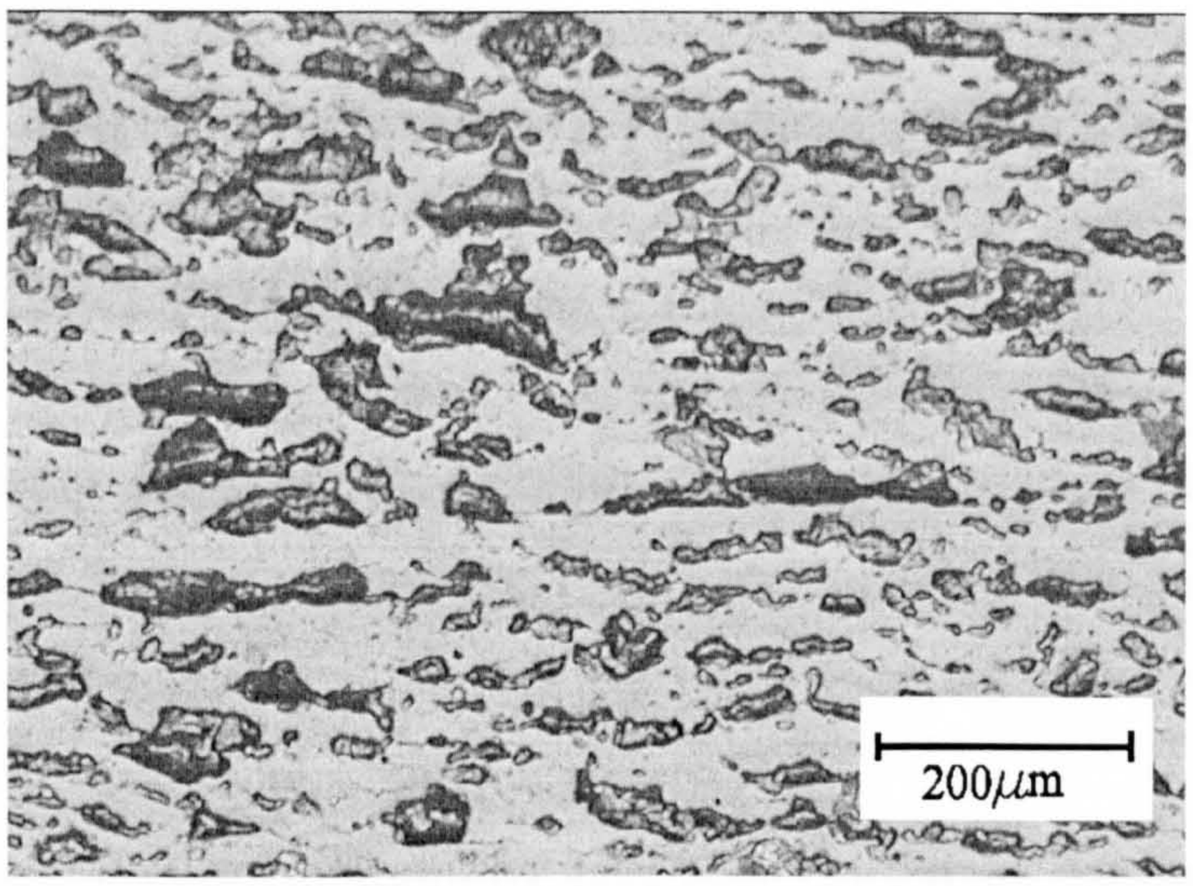


Figure 1-1. Micropitting on the gear tooth flank.

The British Standard Institution (BS 7848: 1996) defines micropitting as “a form of surface fatigue phenomena, which consists of degradation of gear tooth working surfaces under lubrication conditions where the film is too thin for the load”.

Micropitting is increasingly being observed by industry in hardened gears. All types of gears, spur, helical and bevel are susceptible to micropitting regardless the heat treatment applied (carburising, induction hardening, through hardening and nitriding) and the problem occurs with both synthetic and mineral lubricants. Micropitting can occur after a relatively short period of operation, often less than 10^6 cycles.

Gears affected by micropitting need to be replaced due to the increased noise and vibration caused by the deviations of the tooth profile. Continuing operation of the affected gears can lead to tooth breakage, which is a catastrophic type of failure.
The economic and safety implications justify the increasing interest in understanding and preventing micropitting.

1.2 Objectives

The aim of this project was to determine the mechanism of micropitting in gears by assessing the influence of seven controlling factors and by investigating microstructural transformations associated with the micropitting phenomenon.

1.3 A new approach for investigating micropitting

Since the pioneering work of Way (1935) significant progress has been made in understanding the mechanism of macropitting. Macropitting is still observed in clean steels but the evidence indicates that it is nucleated from surface micropits. Currently the mechanism of micropitting is not well understood.
Extensive investigations into micropitting have been carried out previously but a unified methodology considering all the factors involved has not yet been achieved. In order to describe the mechanism it is necessary to identify the factors of influence, to monitor them and to evaluate their effect. In the current work a global approach has been used for studying micropitting by taking into account seven factors that have been previously reported in the literature to influence micropitting. The factors considered are material, surface topography, load, lubricant chemistry, temperature, speed and, slide-to-roll ratio. The assessment of these factors has been done by a fractional factorial experimental design.
The experimental study was carried out using a two-disc machine. The rig has been designed (at the University of Newcastle) to produce contacts that are as realistic as possible with respect to operating conditions in gears.

Several non-destructive (i.e., light microscopy, SEM, optical profilometry, AFM, X-ray diffraction) and destructive (i.e., micro- and nano-indentation) techniques were employed to monitor the surface condition before, during and after testing.

It is well known that in rolling bearings specific microstructural transformations occur below the running surface. They are recognisable due to different etching characteristics when polished specimens are observed in light microscopy. The new phases are known as *dark etching regions*, *white etching bands*, and *butterflies* or sometimes they are referred to as a single term *martensite decay*. Since micropitting is a problem common to bearings and gears the question that arises is whether these transformations occur in gears. In this work extensive metallographic investigations have been carried out on the disc specimens and on helical gears tested at the Design Unit – Gear Technology Centre, University of Newcastle.

1.4 Thesis structure

The complexity of the subject requires a robust theoretical background. Chapter 2 provides information on gear classification (2.2), gear manufacturing (2.3) and heat treatment (2.4). Case carburised steels are treated in a separate section (2.5) in which the microstructure achieved in the case and the core is also presented. Because stresses in gear and surface roughness are highly relevant to the fatigue behaviour these subjects are treated each in a separate section, 2.7 and 2.8 respectively. An important part of Chapter 2 deals with gear fluid lubrication (2.9).

Chapter 3 presents a comprehensive literature review on micropitting and related phenomena (i.e., microstructural transformations). The chapter begins with a classification of the gear failure modes (3.2) which is followed by detailed considerations of the surface contact fatigue phenomena (3.3). The microstructural transformations associated with surface contact fatigue phenomena are described in detail in section 3.4. The characteristics of micropitting, the mechanisms, the factors of influence and the methods for prevention that have been proposed in the literature are presented in sections 3.5 to 3.8.

The experimental methods used to investigate micropitting in disc specimens and in gears are presented in Chapter 4. This chapter introduces the double disc machine used to simulate gear tooth contact (4.2), the statistical approach to design the experiments (4.3), the chemical composition, the mechanical and physical properties of the steels and the grinding methods used to achieve the required surface finish (4.4). It also presents the lubricant properties (4.5) and the operating conditions during tests (4.6). In section 4.7 the techniques used to investigate and quantify micropitting are described.

Chapter 5 presents the results obtained from the investigations carried out during micropitting tests. The results characterise the surface topography (5.2), the wear rate (5.3), the disc diameter (5.4), surface roughness (5.5), the morphology of micropitting (5.6) and microcracks (5.7) as well as residual stress and retained austenite (5.8) and the hardness of the testing surface (5.9).

The results of the metallographic investigations done on disc specimens are presented in Chapter 6. Martensite decay is treated in section 6.2. A new microstructural feature, referred to as the *plastic deformation layer*, has been found in the disc specimens. This is treated in section 6.3

The micropitting results obtained for gears are presented in Chapter 7. The results refer to surface topography (7.2) crack morphology (7.3) and phase transformations (7.4).

Chapter 8 presents the results of the statistical analysis of the factors influencing micropitting. The factors of influence are described in section 8.2. The fractional factorial design has been applied twice. At first in order to assess the influence of the factors on micropitting initiation (8.3) then in order to analyse the influence of the factors on micropitting propagation (8.4).

The conclusions drawn from this research are presented in Chapter 9 together with a brief discussion.

~ Chapter 2 ~

BASIC GEAR THEORY

2 BASIC GEAR THEORY

2.1 General

Since the dawn of engineering transmission systems have been used to transform speed and torque or to change the axis of motion. Essentially, their role is to transmit power and motion from a driving element or a power source to a driven element. The power source can be as rudimentary as muscle power or a natural resource like wind or water or, in the modern times, internal combustion engines, electric motors, steam and gas turbines, etc. The driven device also has a large variety from the wind or water mill to vehicle wheels and many others. There is an impressive number of devices that can be employed (e.g. chains, belts, pulleys) but gears are the most commonly used machine elements for motion and power transmission because of their flexibility, ease of changing speed, direction, torque, compactness, power/weight ratio, etc.

This chapter provides the theoretical background necessary for understanding micropitting and related phenomena.

2.2 Gear classification

A gear is a toothed machine element mounted on a shaft that transmits power and/or motion to a similar element mounted on an adjacent shaft by means of successively engaging teeth. Gears have been used since ancient times. According to Dowson (1979) around 300 B.C. in Alexandria wooden gears were used for toys, catapults, etc. but since the Renaissance, gears have been involved in almost every type of machinery.

The high efficiency of mechanical power transmission (more than 90 %, Snidle et al., 2000) dictates their use for many industrial applications. Gears are nowadays made in a great variety of forms and sizes depending on their application. A first classification can be made according to the relative position of the axes of revolution. From this point of view two categories can be distinguished (Dudley, 1962): gears with parallel axes (see

Figure 2-1) and gears with non-parallel axes (see Figure 2-2). The former category comprises *spur* and *helical gears*. The latter includes *bevel* and *worm* gears.

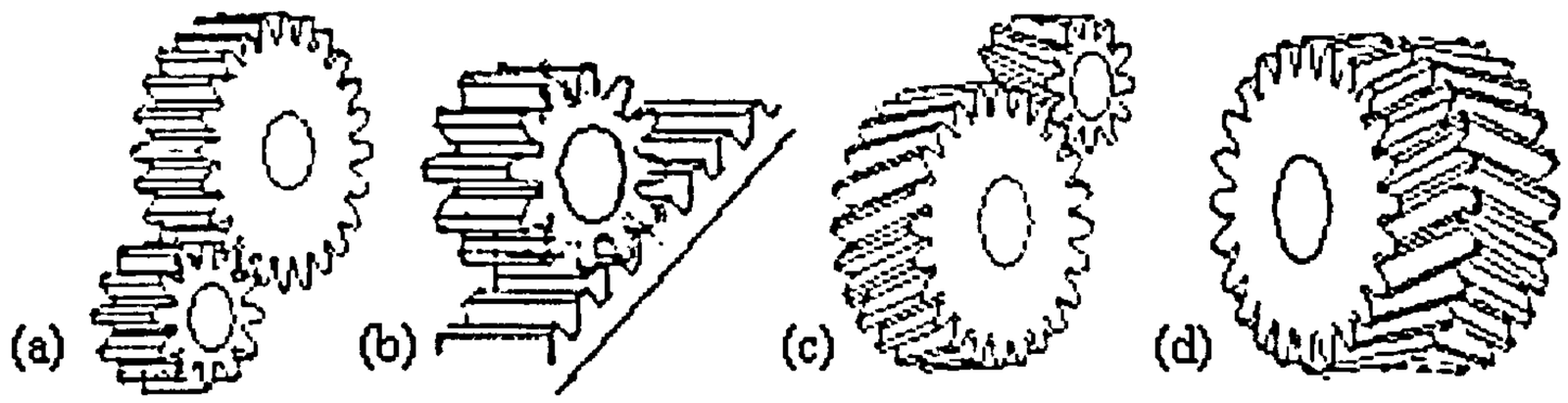


Figure 2-1. Gears with parallel axes: (a) spur gears, (b) rack and pinion, (c) helical gears and, (d) double-helical gears.

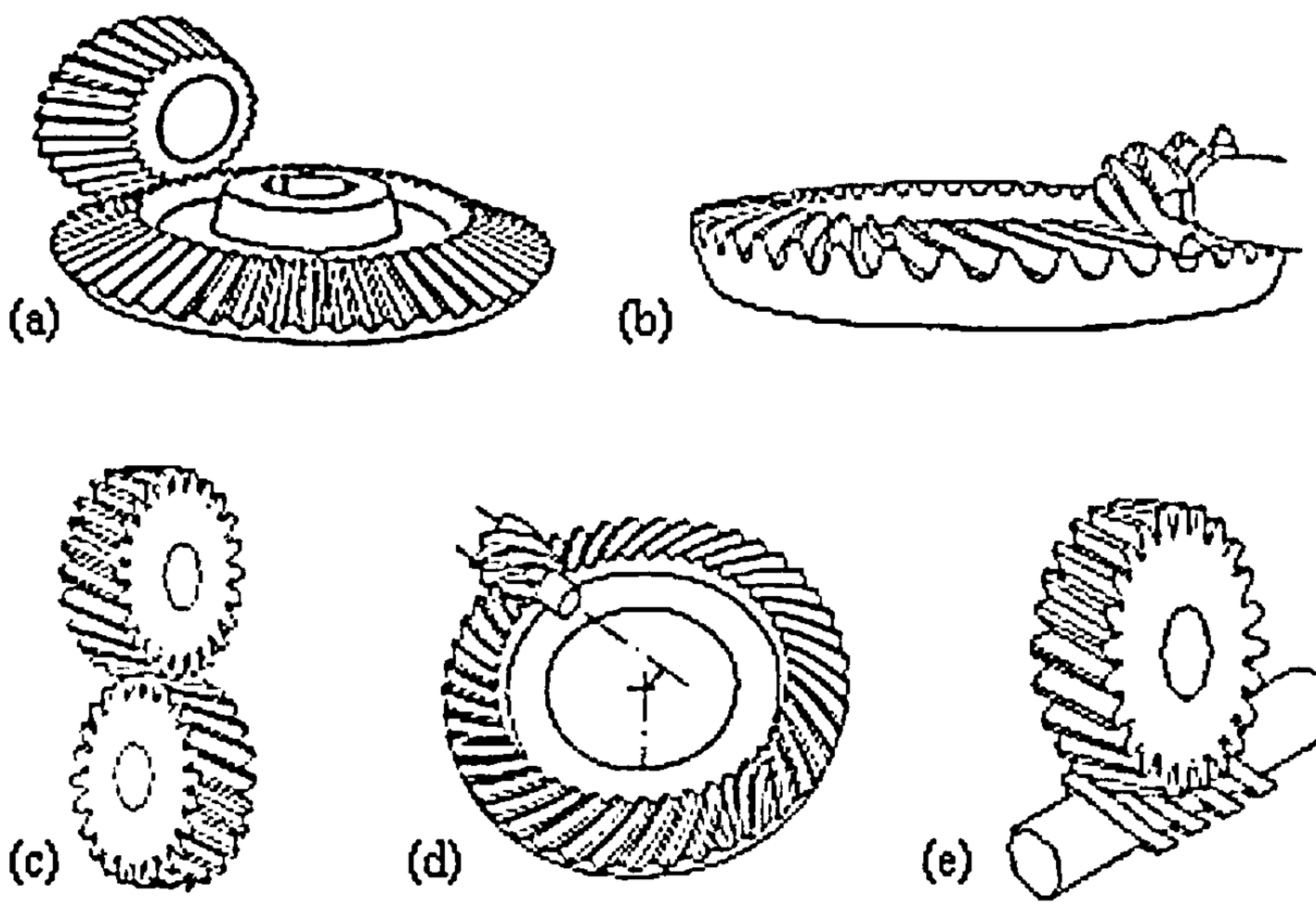


Figure 2-2. Gears with non-parallel axes: (a) straight bevel gears, (b) spiral bevel gears, (c) crossed helical gears, (d) hypoid gears and, (e) worm gears.

Another classification relates to the position of the teeth sides. According to this classification gears can be with teeth that have straight sides and with teeth which have helical sides.

In what follows the most common types of gears are briefly described.

2.2.1 Spur gears

Spur gears are the most popular type representing more than 80 % of gear production (according to Dabnichki & Crocombe, 1999). They are cylindrical and the teeth, which are of involute form, are perpendicular to the face of the gear. Spur gears are used to transmit motion between parallel shafts (Figure 2-1a) or between a shaft and a rack (a gear with effectively an infinite radius, Figure 2-1b), usually, for non-critical applications such as in gearboxes. Despite their advantages (they are the least expensive to manufacture and very efficient) they cannot be used in applications where a direction change between shafts is required.

2.2.2 Helical gears

Helical gears are often used to replace spur gears when higher running speeds are required. The teeth are inclined at an angle to the shaft and they may lie in one (Figure 2-1c) or two bands (Figure 2-1d) around the gear. Helical gear teeth are longer than the teeth on a spur gear of equivalent pitch diameter, which makes them have a greater surface contact than spur gears. This allows a helical gear to carry more load than a spur gear but, at the same time, it reduces the efficiency relative to a spur gear. They can also be made to mesh with non-parallel axes (see Figure 2-2c) but the tooth contact is only at a point, which reduces the load capacity. Their major disadvantage is that they are expensive and less efficient than a spur gear of the same size.

2.2.3 Bevel gears

Bevel gears (Figure 2-2a & b) are used to turn the axis of rotation through 90° where such a configuration is demanded. They are commonly used in a 1:1 ratio just for this purpose. Bevel gears are conical in form and the teeth profiles are of special generated forms (not involute). Straight bevel gears (Fig. 2-2a) have straight tooth elements and spiral bevel gears (Fig. 2-2b) have teeth, which are curved and oblique. One of their most common applications is the bevel gear differential.

2.2.4 Hypoid gears

Hypoid gears (Figure 2-2d) are commonly used for the rear axle of cars, where they provide a right angle speed reducer, with typical ratios in the range 1:3 to 1:6. They usually have teeth that are curved and oblique.

2.2.5 Worm gears

Worm gears (Figure 2-2e) can carry large loads at high speeds. A worm is a gear with teeth in the form of screw threads. They usually work smoothly in a range of ratios from 1:5 to 1:100 but for higher ratios they are not as efficient as spur, helical or bevel gears because of the sliding action inherent in their operation. They are called "self locking" because of the irreversible meshing (only the worm can drive the spur gear and not vice-versa). This can be regarded as an advantage or a disadvantage, depending on the application.

Micropitting, the main subject of this work has been observed to affect mainly spur and helical gears but all gears are susceptible (Erichello, 2000).

2.3 Gear manufacturing background

A large variety of materials is used for gear manufacturing (such as: non-metallic materials, nonferrous alloys, cast iron, powder materials), but steels are the most widely used due to their mechanical properties and relatively low price. The majority of gears are made from forged blanks or rolled stock of carbon and low-alloy steels. There are three different principles that are used for manufacturing gears: *generation*, *form machining* and *moulding* (Hofmann et al., 1991).

The generation principle consists of the synchronised movement of a tool, which is designed as an equivalent rack, with the rotation of the gear blank. The following processes are based on the generation principle: hobbing, shaping, planing, shaving, honing and some grinding methods.

Form machining involves a tool profile that is shaped as an exact replica of the tooth gap to be machined. This principle is used in: milling, broaching, blanking and some grinding methods.

Form moulding is a principle in which gears are produced in one or a sequence of continuous operations. It is achieved by: injection moulding, die casting, sintering and forging.

Gear teeth may be used following the above processes or they may require to be finished by *grinding*, *shaving* or *lapping*.

2.3.1 Hardened gears manufacturing

Hardened gears are most commonly made by one of the following sequences of processes (Hofmann et al., 1991):

(a) Hobbing, shaping or planing followed by shaving and heat treatment;

(b) Pre-finish hobbing, shaping or planing followed by heat treatment and grinding;

(c) Pre-finish hobbing followed by heat treatment and hard hobbing.

Sequence (a) is used for large production quantities when the heat treatment is controlled accurately. Method (b) is used for very accurate gears made in relatively small quantities. The (c) sequence is suitable for high volume production even though it is not as accurate as finish grinding.

2.3.1.1 Hobbing

A hob is a worm like an involute helicoid whose teeth are used as cutting edges. As the hob rotates, the cutting edges sweep out the shape of a rack. The hob rotates in synchronism with the gear blank to cut the required number of teeth (see Figure 2-3).

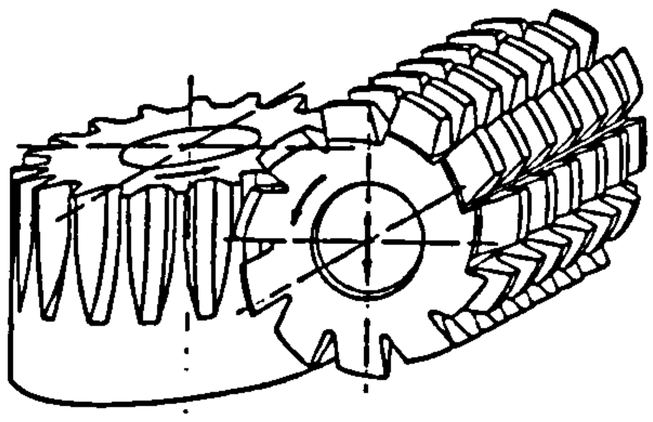


Figure 2-3. Schematic representation of the hobbing process (after Hofmann et al., 1991).

2.3.1.2 Shaping and planing

The cutter (a gear in shaping – see Figure 2-4a and a rack in planing – see Figure 2-4b) moves (rotation or tangentially motion) in synchronism with the gear blank to produce the required number of teeth. Shaping is suitable for internal gears.

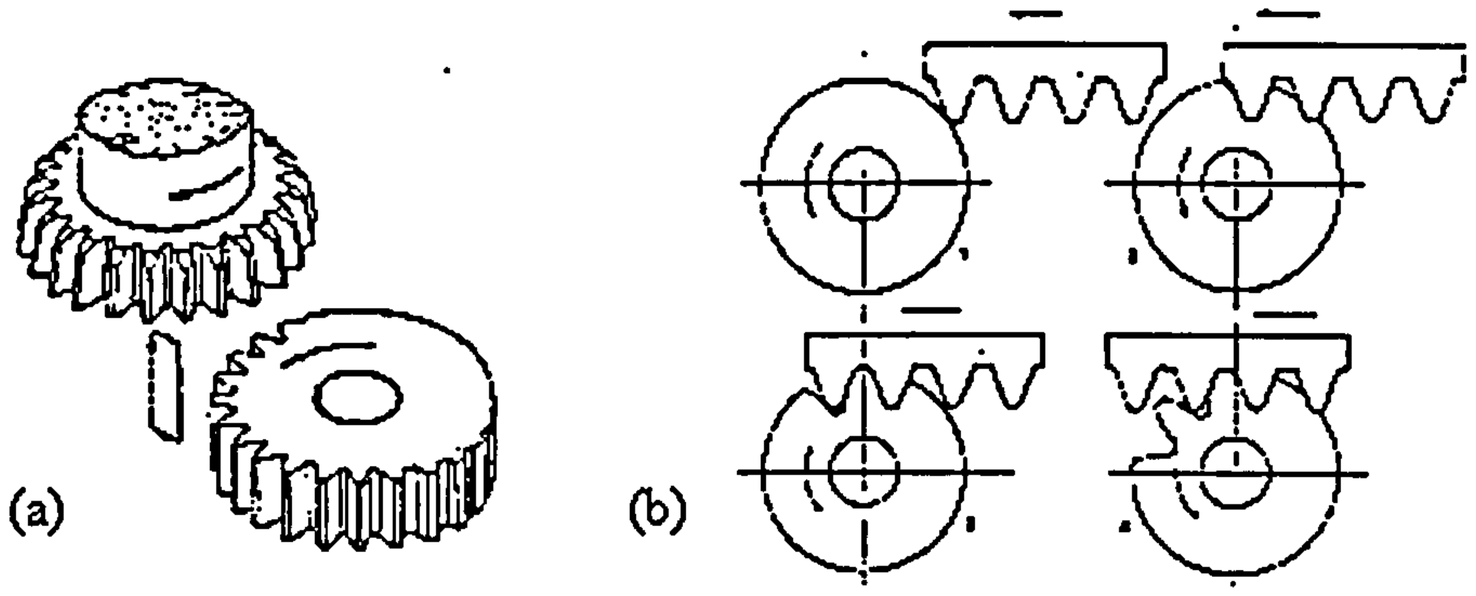


Figure 2-4. Schematic representation of the: (a) shaping process and (b) planing process (after Hofmann et al., 1991).

2.3.1.3 Shaving

Shaving is a free-cutting gear-finishing operation (see Figure 2-5), which removes small amounts of metal from the working surfaces of the gear teeth. It corrects the inaccuracies from gear cutting.

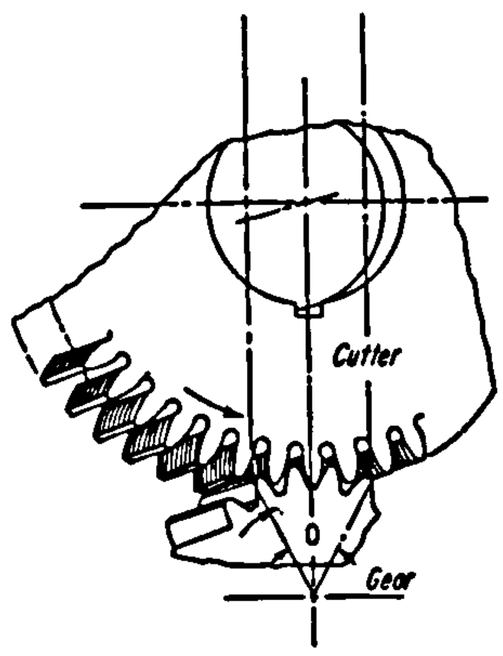


Figure 2-5. Schematic representation of the shaving process (after Dudley, 1962).

2.3.1.4 Grinding

Grinding is most often used to finish gear teeth, but in some instances is used to cut teeth from the solid. The most common grinding methods are (Hofmann et al., 1991):

flat disc (Maag), *involute helicoid* (Reishauer), *single conical wheel* (Höfler & Niles) and *form grinding*.

2.3.1.4.1 Flat disc grinding (Maag)

The grinding wheels are arranged to grind opposite tooth flanks. The wheel faces are either parallel or they form the shape of a basic rack (see Figure 2-6). The gear generates an involute motion by simultaneously rotating and moving tangentially. The process is very accurate even though relatively slow.

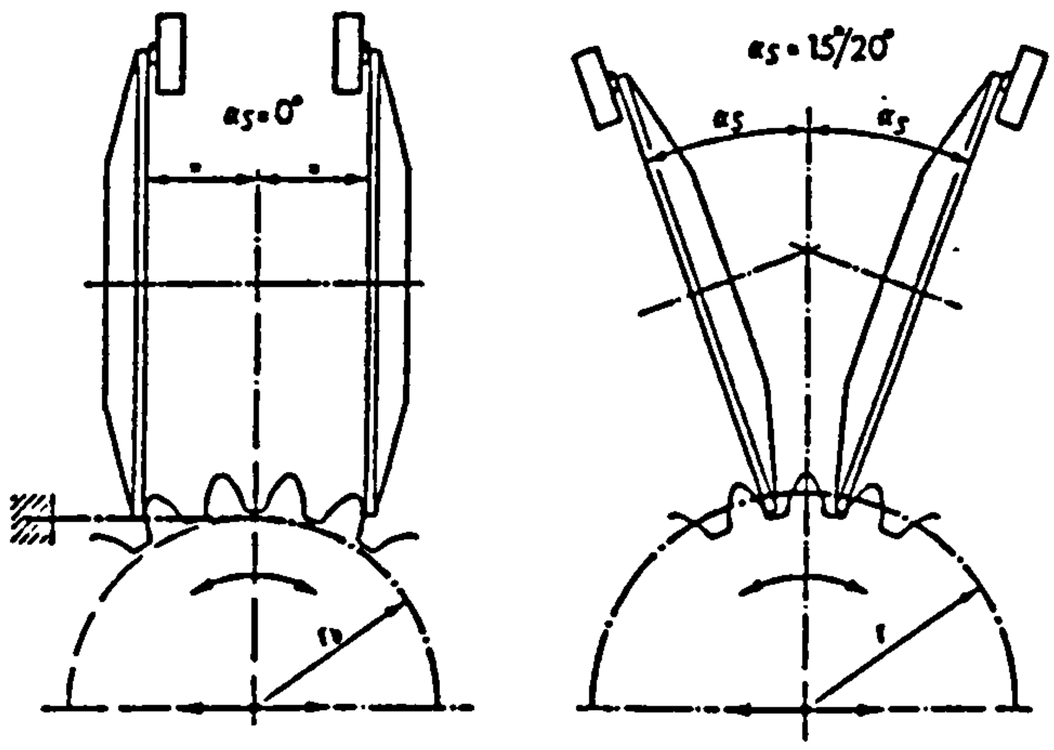


Figure 2-6. Flat disc grinding (after Hofmann et al., 1991)

2.3.1.4.2 Involute helicoid grinding (Reishauer)

The grinding wheel resembles a hob (see Figure 2-7). It is an involute helicoid having a cross section of the required rack form. The motion of the grinding wheel and the gear are synchronised as in the hobbing process. The process is suitable for large volume production being very quick.

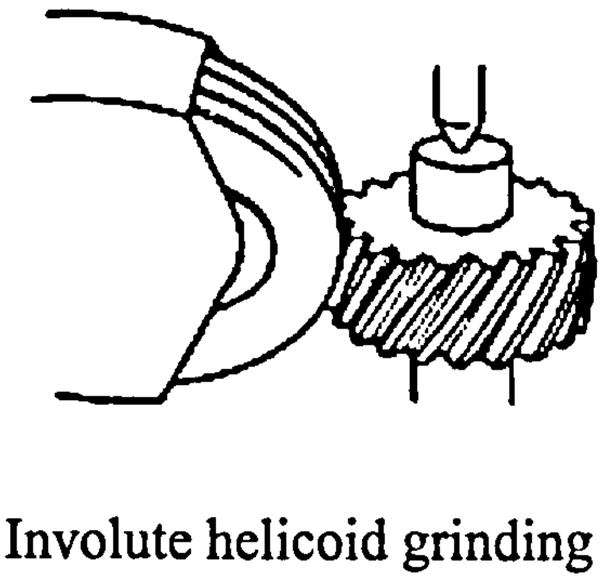


Figure 2-7. Involute helicoid grinding (Reishauer).

2.3.1.4.3 Single conical wheel grinding (Höfler & Niles)

The grinding wheel has the profile of a single rack tooth (see Figure 2-8). The wheel moves tangentially in synchronism with the rotation of the gear to generate the required involute form.

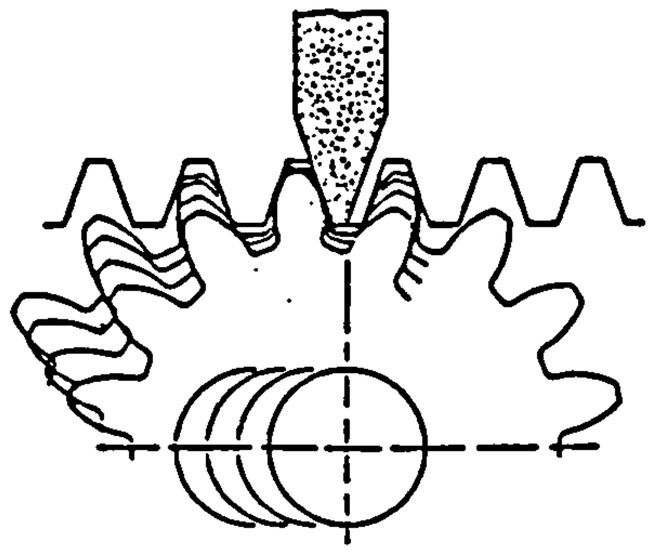


Figure 2-8. Single conical wheel grinding (Höfler & Niles).

2.3.1.4.4 Form grinding

The grinding wheel is used to grind both sides of the space between two adjacent teeth while the gear is maintained in a fixed circumferential position (see Figure 2-9). Because of the high cost of the grinding wheel the process is most suitable for large-scale production. It is also quick and accurate.

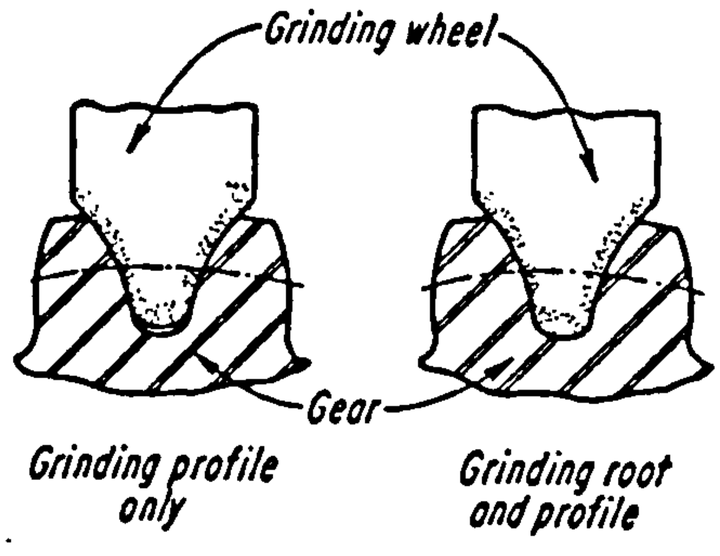


Figure 2-9. Form grinding (after Dudley, 1962).

2.4 Gears heat treatment

A steel used for gear manufacturing must provide, on one hand enhanced ductility in the core to withstand the bending stresses developed under loading conditions and, on the other hand it must have a hard tooth surface in order to resist the shear stresses

developed at and near the surface due to contact (Cavallaro et al., 1995). Since an increase in hardness above a certain level leads to material embrittlement these requirements can be achieved by increasing hardness only in a thin surface layer. For this purpose steels are subject to several heat treatments or thermo-chemical treatments or a combination of both.

The majority of gears that have industrial, automotive or aerospace applications are heat treated either by through-hardening or by case-hardening.

2.4.1 Through hardening

Through hardening is generally applied to gears which do not require high surface hardness. In order to achieve a high hardenability steels used for through-hardened gears are low alloy steels (up to 3 % content of alloying elements) with a medium carbon content (0.3-0.6 %). Through hardening is basically a quench and temper process, which involves the heating of gears in the austenitic domain (see Fe-C phase diagram in Appendix 1) followed by rapid quenching. Gears hardened in this way are then tempered, generally below 200°C. Mechanical properties of the surface are similar to mechanical properties of the core. Through hardened gears are used in gearboxes with large gears that cannot be case hardened.

2.4.2 Case hardening

Case hardening is applied when a gradient of properties (hardness) is desired and it can be done by one of the following processes: *induction hardening*, *carburising and hardening*, *nitriding* or *carbonitriding*,

2.4.2.1 Induction hardening

Induction hardening leads to an increase in surface hardness without altering the chemical composition of the surface layers. Small gears can be rapidly heated by electromagnetic induction produced by a coil that surrounds the gear, then quenched by water jets passing through the inductor coils. Steels with 0.35 to 0.50 % carbon content

are suitable for induction hardening. The case depth depends on the frequency and power density of electrical power and its time duration. In practice the common case depths are about 2 mm. A variant used for larger gears is counter hardening where a shaped inductor is used to heat the material on each side of a single tooth gap.

2.4.2.2 Carburising

Carburising and subsequent hardening provide a hard case and a relatively soft but tough core. Gears are heated above the critical temperature (austenitic domain) in an environment of sufficient carbon potential that allows the absorption of carbon at the surface. Due to carbon diffusion from surface to the interior, a gradient of carbon content is achieved. After carburising, gears are subjected to a hardening treatment. This is done by quenching the gears in a suitable environment (usually oil or water) in order to develop a martensitic or bainitic microstructure in the case, followed by tempering. Tempering consists in reheating gears to a temperature below the transformation range of the steel and holding at this temperature to reduce residual stresses and increase toughness (see section 2.5.4). In practice for gears the case depth is between 0.7 and 1.3 mm but can, in principle, be up to 10 mm.

2.4.2.3 Nitriding

Nitriding is a thermo-chemical treatment frequently used for small precision gears, which are not subjected to high shock loads, or for gears that need to maintain their surface hardness at elevated temperatures. The process, which leads to nitrogen enrichment in a thin surface layer, can be done in gas or liquid media containing nitrogen or, more recently, in a plasma (Rakhit, 2000). The case depth is smaller than for carburising and is usually less than 0.7 mm. Gear steels suitable for nitriding have a carbon content higher than those for carburising and are, generally, aluminium alloyed to allow the formation of aluminium nitrides. The mixture of gamma prime, γ' (Fe_4N) and epsilon, ε ($Fe_{2-x}N$) nitrides that forms at the surface of the case during nitriding is known as the *white layer* due to its appearance in the light microscope (not being etched by nital). It is generally believed the white layer is detrimental to gear fatigue life due to its brittleness but Bull et al. (1999) have shown that the white layer acts like a barrier to

crack propagation, which makes nitrided gears more resistant to micropitting than the carburised ones.

2.4.2.4 Carbonitriding

Carbonitriding is a process in which carbon and alloy steel gears are held at a temperature above the transformation range in a gaseous atmosphere containing carbon and nitrogen. After carbonitriding gears are quenched in order to obtain the desired mechanical properties. The achieved case depth is normally less than 0.7 mm. All steels suitable for carburising can be carbonitrided. For low alloy steels the carbonitrided case has similar properties to that obtained by carburising but the core has lower hardness. For this reason, the gears are more susceptible to quench cracking and carbonitriding is generally applied to gears of low-duty cycle.

2.5 Case carburised steels

Over the past fifty years case carburised steels have been increasingly used for gear applications due to the advantages provided by their combination of properties: surface hardness and core strength. There is a wide range of steel grades that can be carburised but carbon and alloy steels (nickel, chromium, manganese, molybdenum) that contain less than about 0.40 % carbon (usually less than 0.25 % carbon) are best suited for carburising (ASM, 1977). As already mentioned, the aim of carburising and subsequent hardening is to achieve a hard tooth surface and a relatively ductile but tough core. The surface hardness depends on the carbon content at the surface, whereas the core hardness depends on the steel hardenability. *Hardenability* is a material property that can be defined as the ability of steel to increase its hardness during quenching (effectively the ease of martensite formation). Hardenability is increased by the elements dissolved in austenite, coarse grains of austenite and the homogeneity of austenite and is reduced by fine grains of austenite and undissolved inclusions such as carbides or nitrides and non-metallic inclusions (Rakhit, 2000).

The effects of the most significant alloying elements are briefly given below.

Nickel (Ni) increases hardenability by lowering the critical temperatures; therefore martensite can be obtained at lower cooling rates.
Chromium (Cr) increases hardenability and wear resistance. It causes brittleness when steel is tempered above 540°C.
Molybdenum (Mo) induces ductility and toughness and reduces temper embrittlement. It increases hardenability.
Vanadium (V) is usually used together with nickel, chromium and molybdenum. Improves fatigue resistance and acts as an inhibitor for grain growth by carbide formation.

The influence of carbon content and alloying elements on the martensite transformation is presented in the section 2.5.3. Carburised gears are generally heat treated in the following sequence: carburising – quenching – tempering.

2.5.1 Carburising

The process of carburising introduces carbon to the surface of the steel, and the carbon diffuses into the low carbon interior. After carburising a case of high carbon content (up to 1 % carbon) is formed. The high carbon potential environment used for carburising can be solid (pack carburising), liquid (liquid carburising) or gaseous (gas carburising).
Pack carburising is a process in which all surfaces of the workpieces are covered by a solid compound that can provide carbon monoxide when it is heated to a certain temperature. In liquid carburising the source of carbon is a salt bath, with or without molten cyanides (ASM, 1977).

Nowadays, gears are almost exclusively carburised in a gaseous atmosphere. The gaseous atmosphere may be any carrier gas containing nitrogen, carbon monoxide and hydrogen to which a hydrocarbon gas has been added. The most commonly used is an endothermic gas produced by the reaction of methane with oxygen.
The number of chemical reactions taking place simultaneously in a gas-carburising atmosphere is estimated to be as many as 180 (ASM, 1977). However, the principal reactions, which determine the carbon transfer to the steel are as follows:

$$Fe + 2CO \rightarrow Fe(C) + CO_2 \quad (2\text{-}1)$$

$$Fe + CH_4 \rightarrow Fe(C) + 2H_2 \quad (2\text{-}2)$$

$$Fe + CO + H_2 \rightarrow Fe(C) + H_2O \quad (2\text{-}3)$$

where Fe(C) represents carbon in solution in austenite.

Free carbon resulting from chemical reaction is adsorbed at the surface due to the gradient carbon potential between atmosphere and steel. The process takes place at a temperature above the critical temperature Ac_3 (see Fe-C phase diagram in Appendix 1) of the steel (usually between 850 and 950°C) and the carbon atoms adsorbed onto the surface set up a concentration gradient, which determines their diffusion from the surface to interior. The amount of carbon and the penetration depth depend on the carbon potential of the environment, the carburising temperature and holding time but cannot exceed the solubility of carbon in austenite.

2.5.2 Quenching

After carburising, gears are quenched (cooled at a rate which suppresses the equilibrium transformations) in a suitable medium, which is commonly oil or water. Because of the cooling rate the diffusion of carbon atoms cannot take place and the austenite transforms without diffusion into a phase called *martensite*. Martensite is a supersaturated solid solution of carbon in iron, which has a body-centred tetragonal (BCT) lattice. Quenching gives rise to a case of high hardness (martensitic) overlying a core of lower hardness (bainitic, martensitic or ferritic). The *depth of case* is defined (Thelning, 1975) as "the distance from the surface to a plane at which the hardness is H_v=550 VHN", where H_v is the Vickers microhardness determined for 300 gf load. Tests carried out by Cram (1961) on carburised and induction-hardened steels indicated that the case depth should be at least twice the depth of the point of maximum shear stress generated by contact in order to prevent high local stress concentrations, which may lead to fatigue crack initiation. The carbon gradient influences martensitic transformation, the morphology of the martensite, and ultimately the mechanical properties of the steel.

2.5.3 Martensitic transformation

According to the iron-carbon phase diagram (see Appendix 1) at 723°C iron changes its crystallographic structure from FCC γ-Fe (austenite) to BCC α-Fe (ferrite). The solubility of carbon in austenite at 723 °C is 0.77 % and only 0.0218 % in ferrite. If steel is cooled down slowly the excess carbon is forced to leave the iron structure by diffusion and austenite, either transforms into pearlite (a mixture of ferrite and cementite) or, decomposes to ferrite and carbide precipitates. When the steel is quenched rapidly from the austenitic domain, there is not enough time for carbon atom diffusion to take place and they remain "trapped" in iron solid solution and only a distortion of the crystal structure takes place (from FCC to BCT). The transformation is called *diffusionless* because individual atomic movements are less than one interatomic distance and is called a *military transformation* because of the collective movement of atoms. Since the transformation free energy needs to be redistributed the result is a shear process, which leads to the formation of needles or plates throughout the parent austenite grain. Nishiyama (1978) has proposed the term "*domino*" because the lenticular units form one after another as in a domino toppling game. The whole process is autocatalytic and it takes place at the speed of sound in the steel (Porter and Easterling, 1992). The first suggestion of a shear transformation mechanism was made by Bain in 1924. He suggested that the transformation involves a 20 % compression of the *c* axis of the austenite unit cell and 12 % expansion of the *a* axis (Figure 2-10).

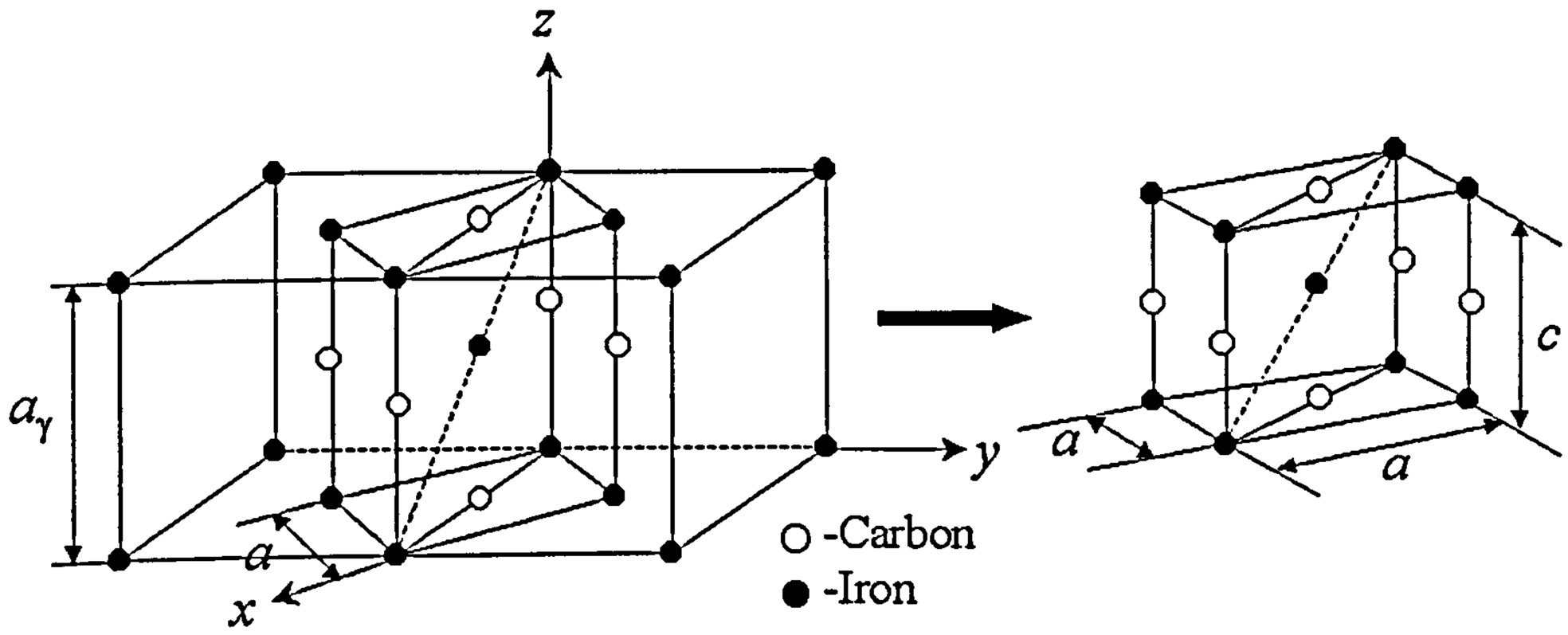


Figure 2-10. Bain correspondence. Austenite unit cell is contracted 20 % on the *c* axis and expanded 12 % on the *a* axis.

Further additions to this theory have been made by Kurdjumov & Sachs, Nishiyama & Wasserman, Greninger & Troiano and Bowles as cited by Smallman 1962 and Porter & Easterling 1992. The *c/a* ratio is given by the equation below (Porter and Easterling, 1992).

$$\frac{c}{a} = 1.005 + 0.045(\text{wt\%C}) \qquad (2\text{-}4)$$

Equation (2-4) shows that the tetragonality of the martensite increases with increasing carbon content.

Martensitic transformation starts at a temperature M_s that depends on the chemical composition according to the Steven-Haynes (Steven & Haynes, 1956) equation (2-5) and finishes at a temperature M_f below which further cooling does not increase the amount of martensite:

$$M_s(°\text{C}) = 561 - 474(\%\text{C}) - 33(\%\text{Mn}) - 17(\%\text{Ni}) - 17(\%\text{Cr}) - 21(\%\text{Mo}) \qquad (2\text{-}5)$$

In practice M_f may not correspond to 100 % martensite and some retained austenite can be left. 10-15 % of retained austenite is a common feature in carburised and quenched steels. M_f is often below room temperature and cryogenic treatment is necessary to get almost 100 % martensite.

2.5.4 Tempering

Martensite is a very hard phase but is normally very brittle. The heat treatment by which the mechanical properties of martensite are modified to increase ductility is called tempering. Tempering consists of heating the steel to a temperature in the range of 150-700°C (Honeycombe, 1981) below the critical temperature, holding it a certain time and cooling it with a rate that allows the equilibrium transformations to develop. The aim of tempering is to enhance ductility without significantly affecting hardness. The changes that occur during tempering depend on temperature and holding time at this temperature. The hardness of tempered martensite decreases due to the growth and coalescence of cementite particles. Gears are commonly tempered in the range 115-175°C (Rakhit, 2000). Martensite is a thermodynamically unstable supersaturated phase with a high density of lattice defects. By increasing temperature the mobility of atoms is increased and diffusion phenomena cause the decomposition of martensite and other

non-equilibrium phases (retained austenite). Depending on the temperature range, tempering process can be classified in three stages.

2.5.4.1 First stage of tempering (150-200°C)

The martensite lattice is distorted by the presence of carbon atoms in excess. As a result of increasing temperature, carbon diffuses and segregates to low energy sites such as lattice defects (dislocations, vacancies, grain boundaries) in order to lower its chemical potential (Leslie, 1981). In twin martensite, which forms in the case, the density of dislocations is reduced and carbon atoms segregate as clusters. Clustering is followed by the precipitation of very fine carbides (~3 to 5 nm according to Leslie, 1981, 2 nm according to Krauss, 1999). These transition carbides are: epsilon (ε) ($Fe_{2.4}C$ with hexagonal lattice) or eta (η) (Fe_2C with monoclinic lattice) (Porter and Easterling, 1992). According to Leslie, 1981, eta (η) carbides do not form at carbon concentrations above 0.8 % but only epsilon (ε) carbides. The precipitation of ε-carbides causes the martensite needles to become dark etching and there is a partial loss of tetragonality in martensite (Schrader & Rose, 1966). The ε-carbides form short rods, which are arranged parallel to each other in groups throughout the whole of the martensite needles (see Figure 2-11).

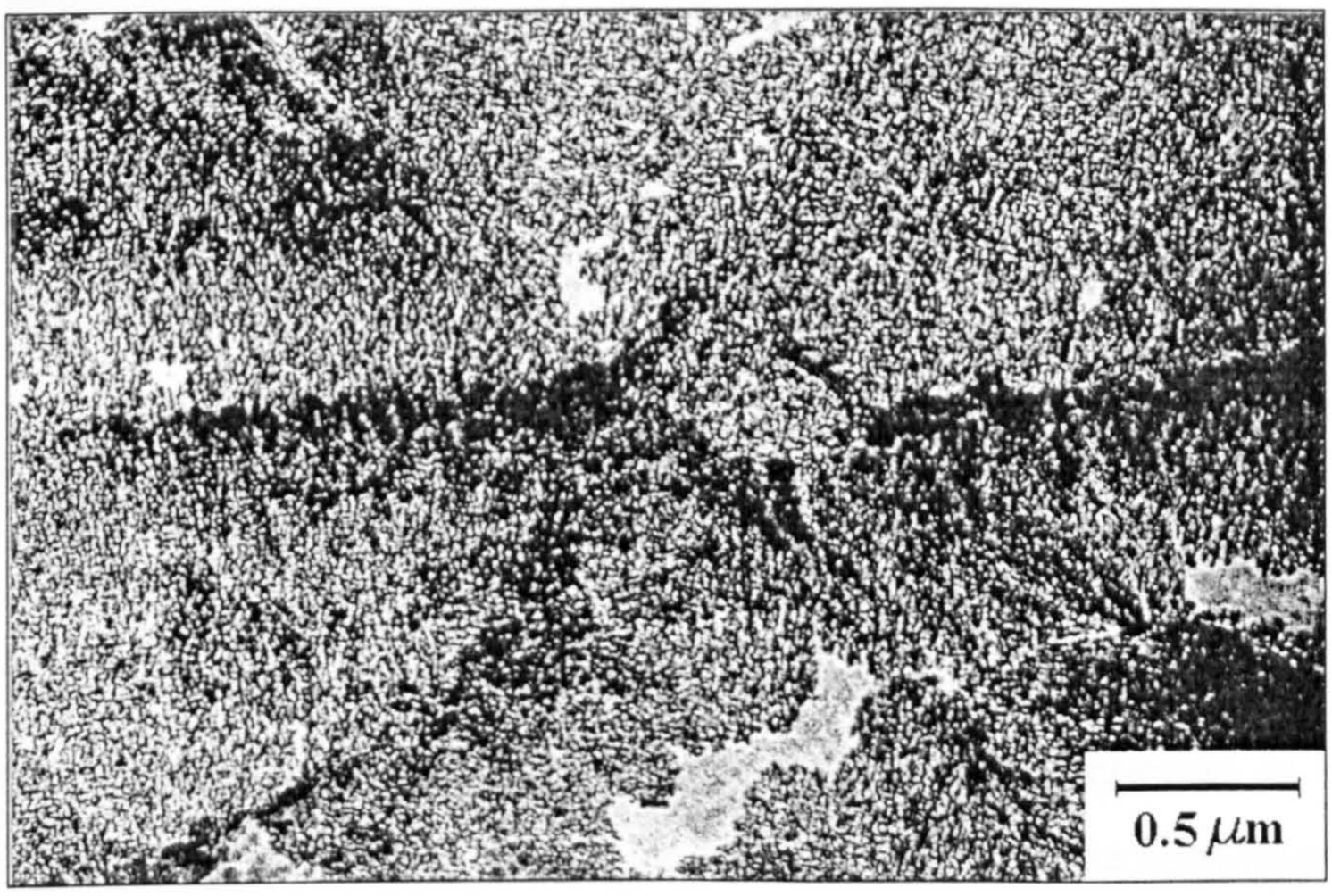


Figure 2-11. First stage of tempering. TEM replica. The ε-carbides form a network of fine rods arranged perpendicular to each other. They have different orientation in different needles. The white areas are retained austenite (from Schrader & Rose, 1966).

2.5.4.2 Second stage of tempering (200-350°C)

This stage is dominated by the transformation of retained austenite to bainite (very fine precipitates of cementite in a ferrite matrix) and the precipitation of cementite from martensite, as shown in Figure 2-12. Cementite precipitation is accompanied by dissolution of the ε-carbide phase so that martensite loses its remaining tetragonality and becomes BCC ferrite. Consequently, hardness decreases to a value corresponding to ferrite. The decomposition of retained austenite starts from the boundaries of the martensite needles, and bainite often grows acicularly into the martensite. The resulting mechanical properties are not suitable for gear applications.

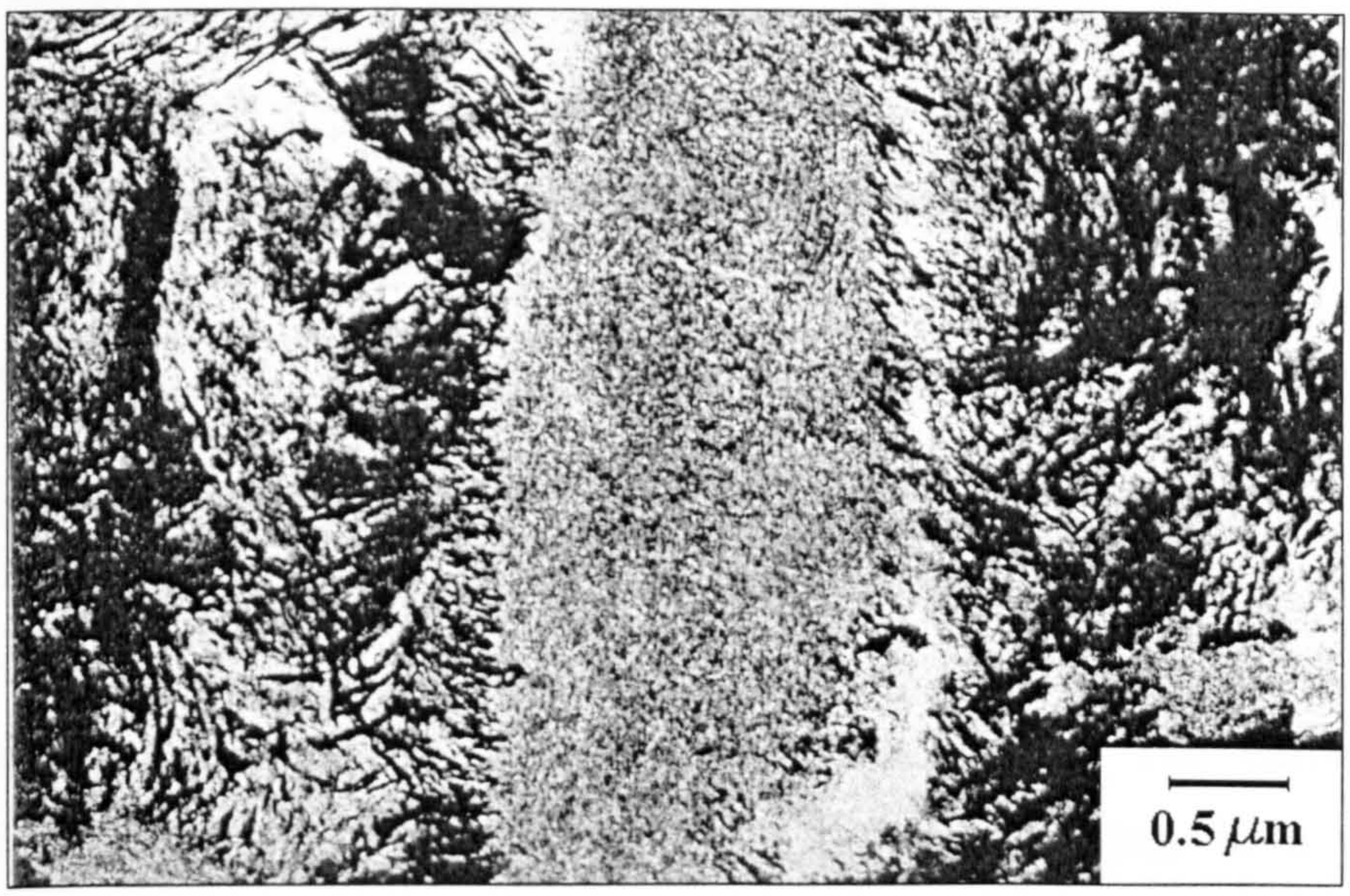


Figure 2-12. Second stage of tempering. TEM replica. Fine precipitates of ε-carbides can be seen within the martensite needle (centre of the picture). Outside the needle coarse cementite is observed. (from Schrader & Rose, 1966).

2.5.4.3 Third stage of tempering (350-600°C)

During the third stage of tempering cementite crystals coarsen in the original martensitic areas and ferrite recrystalises. With increasing temperature in the third stage, growth of the cementite precipitates occurs (see Figure 2-13). After tempering at 600°C, the cementite is coarsened so much that it can be easily resolved in the light microscope.

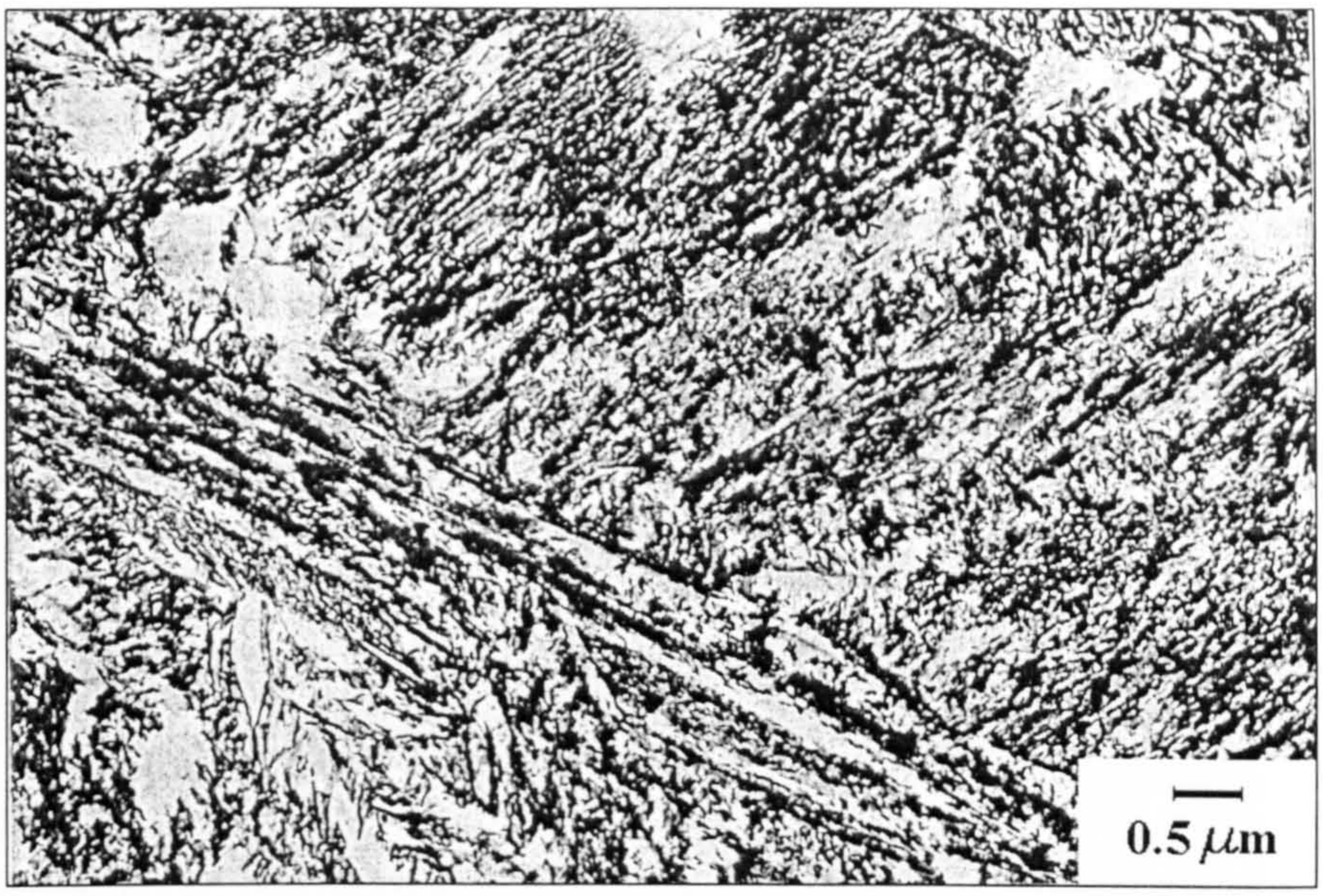


Figure 2-13. Third stage of tempering. TEM replica. Cementite forms small rods, granules or elongated rods (from Schrader & Rose, 1966).

2.5.5 Microstructure of quenched and tempered case carburised steels

The microstructure that can be formed in case and core depends on both, chemical composition of the respective layers and the section size. The carbon gradient introduced in austenite by carburising results, after quenching, in a gradient of martensitic microstructures. The distinction between case and core is made by the case depth, previously defined.

2.5.5.1 Case microstructure

The case microstructure may consist of tempered martensite, retained austenite, massive carbides, non-metallic inclusions, phosphorus segregation, and surface oxides (Krauss, 1995a). The martensite that forms at low M_s (martensite transformation start) temperatures appears in plates, and hence is called *plate martensite*. Plates of martensite are non-parallel and appear under the light microscope like a zigzag array of needles or acicular shapes. The plates etch dark and are surrounded by white-etching zones of retained austenite. A very important feature that occurs during tempering is the development of a high *density of dislocations* (Krauss, 1995a) produced by the

deformations accompanying the martensitic and subsequent tempering transformations. The carbide arrays, and the associated dislocation structures, make deformation by slip or dislocation motion difficult. As a result, the hardness and strength of martensite increases with increasing carbon content. The ductility and toughness of martensitic structures are quite low but it is not completely brittle since some dislocation motion is still possible. According to Krauss (1999), brittle fracture that occurs in the case is attributed to other microstructural features than martensite.

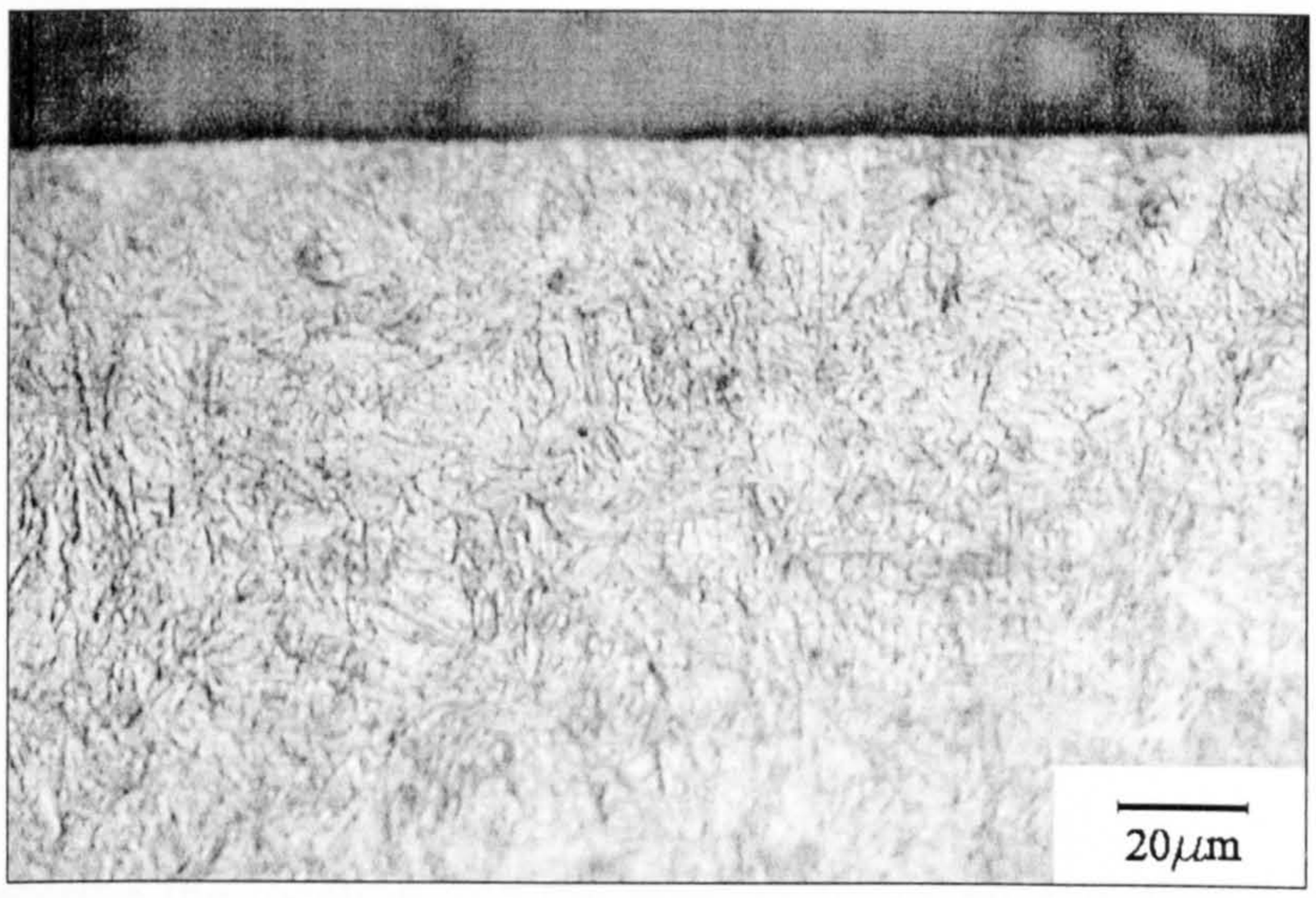


Figure 2-14. Case microstructure of En36 steel consist of tempered martensite.

In practice, gears are tempered below 200°C and, as described above, the only microstructural change produced by tempering is the *precipitation of transition carbides*.

2.5.5.2 Core microstructure

The core microstructure may consist of tempered martensite, bainite or ferrite and pearlite (Krauss, 1995a). The martensite crystals appear to have lath or board-shape geometry and are called *lath martensite*. The crystals are relatively thin and flat with one long dimension and adjacent crystals form parallel to each other in stacks or packets. Most of the crystals are too thin to be resolved in the light microscope. The larger the carbon content, the larger is the amount of retained austenite. The austenite is retained as thin films between the laths and is not resolvable in the light microscope.

Figure 2-15. Core microstructure of EN36 steel consists of a mixture of martensite and bainite. The prior austenite grains are distinguishable.

2.6 Geometric characteristics of spur gears

In order to describe the location and the mechanism of micropitting it is necessary to consider a few geometrical characteristics of gears. Because of their simplicity, spur gears will be used as an example. Gears can be represented kinematically (Dudley, 1962) by *pitch surfaces*, which are cylinders for gears with parallel axes (spur and helical) and, cones for gears with intersecting axes (bevel gears) (see Figure 2-16).

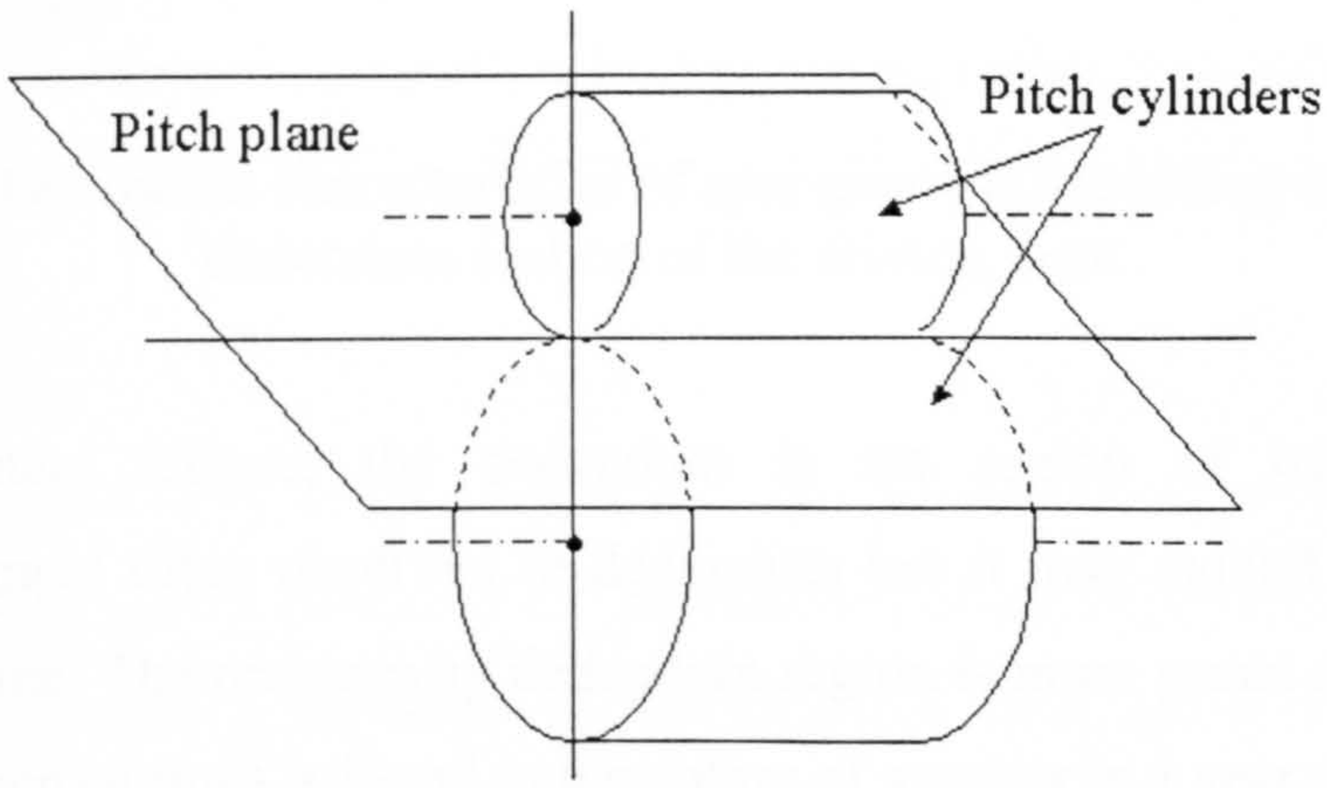


Figure 2-16. Pitch surfaces are cylinders for spur and helical gears.

The pitch plane is mutually tangent to the pitch surfaces and gear teeth are positioned around the pitch surface. The curve most commonly used to generate the tooth profile is the *involute* of a circle, which is defined as the locus of a point on an in-extendable cord, unwound from that circle. The circle from which the involute is developed is called the *base circle* of the involute. Gears have a datum circle called the *reference circle* or *pitch circle*. It is the circle on which the cutter reference line rolls without slip during generation and represents a cross section of the pitch surface.

The main geometric characteristics of a gear tooth are illustrated in Figure 2-17. The circle bounding the tips of the teeth, in a cross section of the gear is the *addendum circle* and the circle bounding the spaces between the teeth, in a cross section of the gear is called the *dedendum circle*. The radial extent of the teeth outside the pitch circle is called *addendum*, and the extent by which the innermost part of the tooth root lies inside the pitch circle is called *dedendum*.

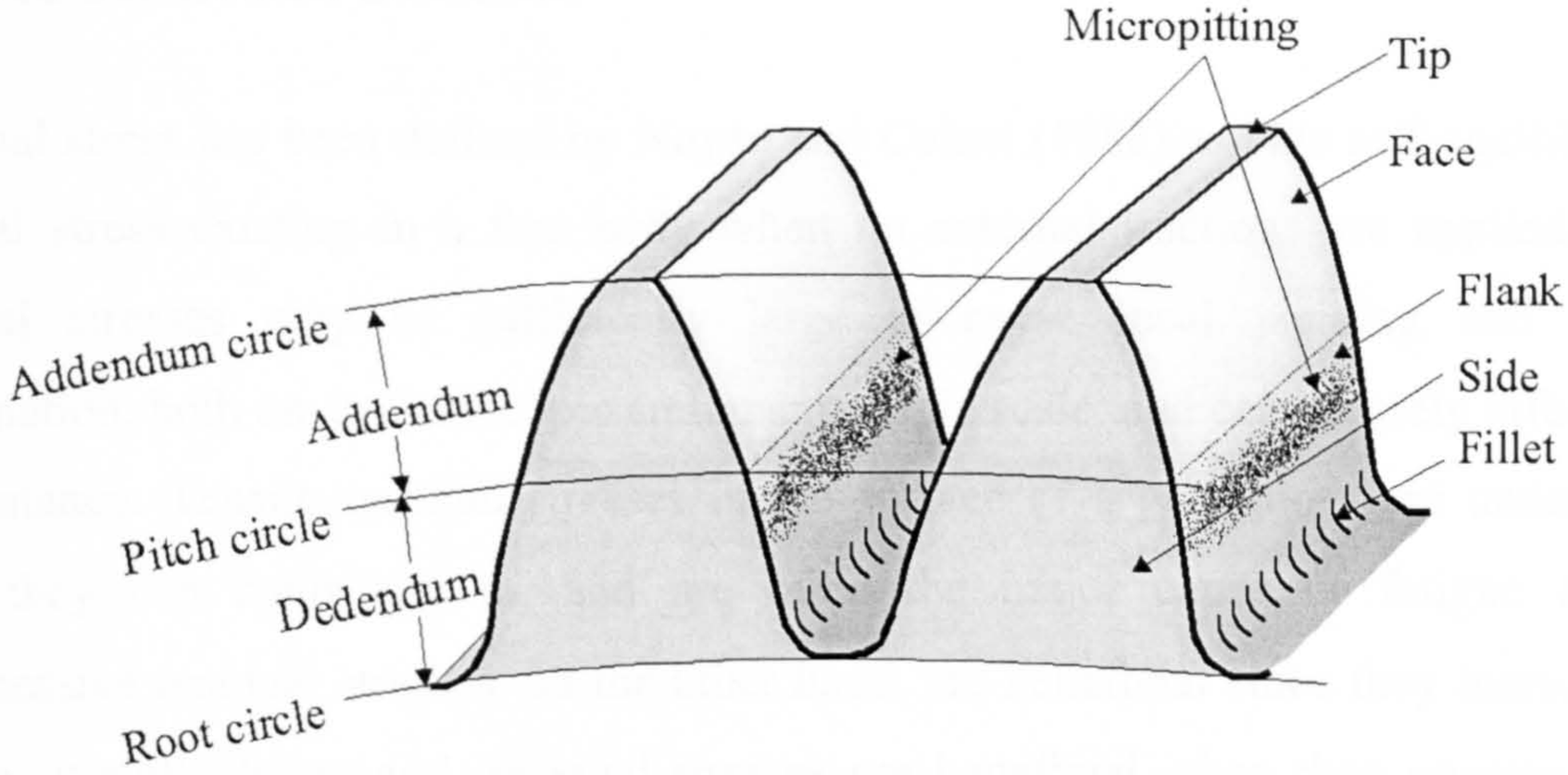


Figure 2-17. Geometric characteristics of spur gears; micropitting initiates in the dedendum section of the driving gear.

Concerning contact fatigue, the dedendum is the region of particular interest. Micropitting is most often observed in dedendum but it may extend in addendum or along the pitch line. The reason why dedendum region is more prone to contact fatigue damages (i.e., micropitting) is found in the nature of stresses in a gear tooth contact (see next section).

2.7 Stresses in gears

The teeth represent the critical parts of a gear. They can fail under working conditions either by *root breakage* or by *surface contact fatigue*. Tooth breakage occurs when the bending stresses exceed the bending strength of the gear steel. Surface failure occurs when the *contact stresses* exceed the fatigue strength of the steel. The fatigue strength depends on the gear steel and the manufacturing processes. In gear design, the reliability during the working life is a very important issue. Stress analysis must therefore account for the stresses existing in gear before any load has been applied. These stresses, known as *residual stresses* have their origin mainly in manufacturing processes. The aim of this section is to provide a good understanding of the stress pattern in gears.

2.7.1 Residual stresses

Residual stress has been defined by Noyan and Cohen (1982) as "the self-equilibrating internal stress existing in a free body when no external tractions are applied." The residual stresses may be sufficiently large to cause local yielding and plastic deformation, both on a microscopic and macroscopic scale, and can severely affect gear performance. Tensile residual stresses in the surface of a gear tooth are undesirable since they can contribute to, and are often the major cause of fatigue failure. Compressive residual stresses, on the other hand, are beneficial since they increase the fatigue strength. In general, residual stresses are beneficial when they operate in the plane of the applied load and are opposite in sense (Kandil et al., 2001).

Residual stresses are produced in a solid body by its inhomogeneous deformation due to thermal gradients and to phase transformation dilatations. Residual stresses present in case carburised steels have their origin in the heat treatments applied and are classified as *thermal stresses* and *transformation stresses*, and they also can be induced by machining (i.e., grinding), hence they are called *mechanical stresses*.

2.7.1.1 Thermal stresses

When the quenching begins the core has a higher temperature than the case. Thus, the core has a greater specific volume than the case. Due to this difference compressive stresses will arise in the core and tensile stresses will arise in the case. If the stresses exceed the yield stress, plastic deformation occurs in the case. Due to plastic deformation the specific volume in the case will become greater than the specific volume in the core. When the difference of temperature between core and case decreases the stresses will be reversed resulting in *compressive residual stresses* in the case and tensile residual stresses in the core (Ericsson and Hildenwall, 1982). Additional thermal stresses are caused by the different alloying elements with different specific volumes. We can conclude that quenching induces *compressive residual stress* in the surface, balanced by tensile residual stresses in the bulk.

2.7.1.2 Transformation stresses

Transformation stresses occur mainly due to the transformation of austenite to martensite, which is accompanied by a volume expansion. During carburising the volume increases due to ferrite to austenite transformation and due to further increase of temperature. During quenching the volume decreases as the austenite cools down and increases again when the transformation of austenite to martensite occurs.

At the beginning of quenching only the surface region transforms to martensite and the region expands. As a result, the inner austenite deforms plastically to accommodate this change. When the interior transforms to martensite its volume expands. Thus, at the end of the quenching *tensile residual stresses* are present in the case.

Measurements done by Ericsson & Hildenwall have shown that the peak of residual stresses corresponds to a certain carbon content which is associated with the maximum amount of transformed martensite. Tensile residual stresses can contribute to crack initiation and propagation. The first stage of tempering eliminates some of these residual stresses (e.g., martensite loses tetragonality due to carbide precipitation).

2.7.1.3 Residual stresses produced by machining

Mechanically generated residual stresses are a result of machining processes that produce non-uniform plastic deformation. During machining (e.g. grinding) the near surface region of the gear tooth is deformed plastically and the volume elements are extended in this region. The constraints of the bulk introduce compressive stresses in the subsurface region (Döle & Cohen, 1980). If work hardening occurs only near the surface during machining, this produces tensile residual stress due to elastic relaxation.

2.7.1.4 Shot-peening process

It is generally agreed that surface *compressive residual stresses* can increase the gear life by delaying the crack initiation and propagation. Fatigue occurs by the progressive growth of surface cracks under cyclic loading. It follows that if the stress at the crack tip is below that necessary for propagation then fatigue failure will not occur. The effect of compressive residual stresses in contact fatigue is usually beneficial and a process that induces such stresses in gear surfaces has been developed recently. Known as *shot peening* this process is a method of cold working in which a high velocity stream of shot causes local plastic yielding in the surface. Due to interior constriction, high, local compressive stresses are induced in the exposed surface layers. The magnitude of compressive stress induced by shot peening can be as high as 60 % of the ultimate tensile stress of the material being treated (Dölle & Cohen, 1980). Recently, Metal Improvement Company (UK) and Lawrence Livermore National Laboratory of the University of California have successfully developed a new method, laser shot peening (Hammersley et al., 2000). The residual stresses induced by laser technology are comparable to those achieved by conventional shot peening, but at much greater depths.

2.7.2 Gear tooth contact

In order to understand how gear teeth can fail, consideration of the gear tooth contact stresses is necessary. The teeth of two contacting spur gears having the radii of curvature R_1 and R_2 and the base circles centred in C_1 and C_2 respectively, can be represented as two equivalent cylinders of diameter $2R_1$ and $2R_2$ as shown in Figure 2-

18. The line O_1O_2 that is tangent to both base circles is called the *line of action*. The instantaneous point of contact P is always situated on the line of action in proper meshing conditions. The line C_1C_2 that connects the centres of the base circles is called the *line of centres* and the distance C_1C_2 is known as *centre distance*. The angle ψ measured at the centre of the base circle (C_1 or C_2) between the line of centres C_1C_2 and the perpendicular to the line of action (C_1O_1 or C_2O_2) is called the *pressure angle*. The point Z at the intersection between the line of centres C_1C_2 and the line of action O_1O_2 is the *pitch point*. At this point there is no relative motion between the two teeth hence the motion is *pure rolling*. At any other point along the line of action sliding occurs and the motion is a combination of *rolling and sliding*.

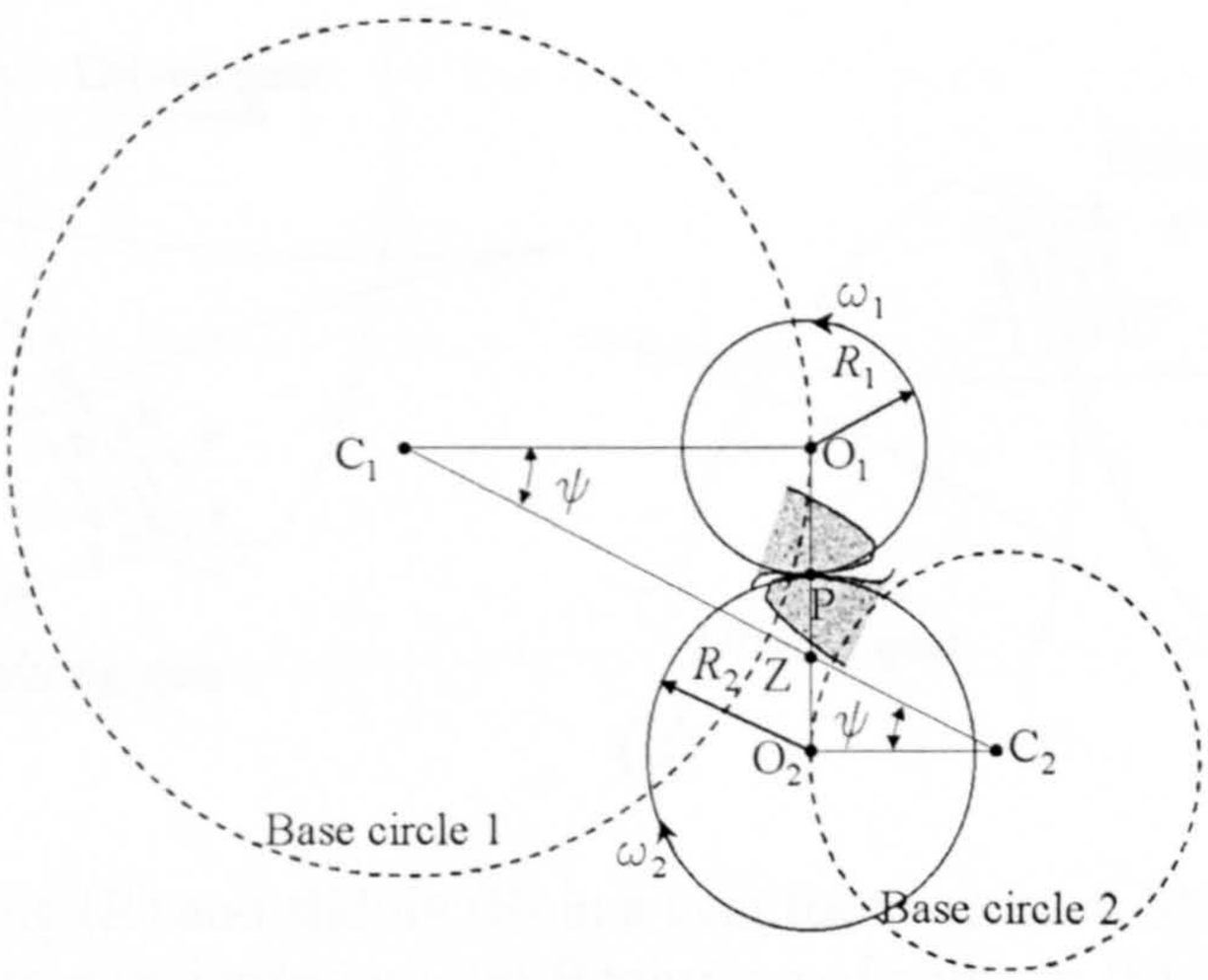


Figure 2-18. Contact between two gear teeth can be described by the contact between the disc centered in O_1 of radius R_1 and the disc centered in O_2 of radius R_2.

The slide-to-roll ratio is given by the Equation 2-6, where u_1, u_2 are the linear velocities of the two surfaces and R_1, R_2 the radii of the equivalent cylinders.

$$S = \frac{2(u_1 - u_2)}{u_1 + u_2} = \frac{2(R_1 - R_2)}{R_1 + R_2} \tag{2-6}$$

If ω_1 and ω_2 are the angular speeds of the two gears then the sliding speed is given by the equation below.

$$v = (\omega_1 + \omega_2) \cdot s \tag{2-7}$$

where, s is the distance of the instantaneous point of contact P from the pitch point Z.

During meshing, gear tooth flanks slide and roll one over another. Along the pitch line the motion is pure rolling but in the addendum and dedendum the motion becomes a combination of rolling and sliding (see Figure 2-19). The contact point of a gear tooth moves opposite to the direction of rotation of the gears. This is the direction of rolling. In the addendum, the direction of sliding and rolling is the same hence the sliding is called *positive*. In the dedendum, rolling and sliding occur in opposite directions hence the sliding is called *negative*. The effect of the negative sliding is that the material at the surface is rolled in one direction and pushed in another resulting higher stresses than those in positive sliding (Fernandes & McDulling, 1997).

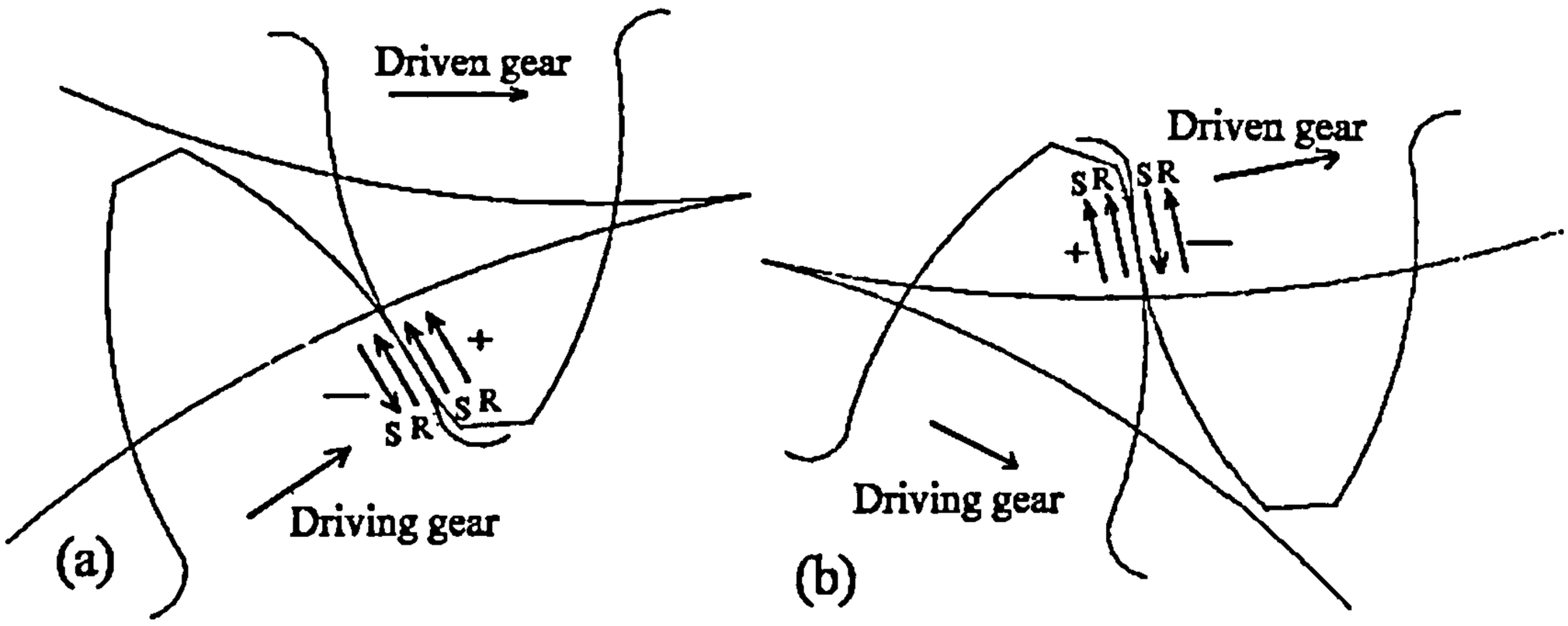


Figure 2-19. Rolling (R) and sliding (S) in a gear tooth contact. Sliding is positive in addendum and negative in dedendum. (a) Beginning of contact; (b) End of contact (after Fernandes & McDuling, 1997)

Sliding also causes heat generation, resulting in a substantial increase of the temperature of the contacting surfaces. Therefore, the dedendum region is more prone to fatigue damage than addendum. In practice, micropitting always initiates in dedendum region of the smaller gear, which is usually the driving gear. This complex nature of motion generates two types of stresses, i.e. *bending stresses* at the root of the tooth and *contact stresses* (surface and sub-surface shear stresses) which arise from the contacting surfaces. Gears are subject to *normal stresses* due to rolling and to *tangential stresses* due to sliding.

The basic kinematics of involute gears gives rise to high contact pressure, which can be as high as 3-4 GPa (Höglund, 1999). Failure of the gear tooth can occur in service when bending stresses or contact stresses exceed the safety limits.

Micropitting and other surface contact fatigue failures occur due to contact stresses. In the next sections contact stresses are analysed in detail while bending stresses are discussed briefly.

2.7.3 Stresses due to loading

Under operating conditions gears are subjected to two types of stresses: bending stresses at the root of the tooth and contact stresses arising from the contacting surfaces.

2.7.3.1 Bending stresses

In spur and helical gears the maximum bending stress occurs at or close to the 30° tangent to the root. In the root area, the rapidly changing section gives rise to stress concentrators. The basic stress formulas can be summarised in general terms (Shipley, 1974) with the help of three indices: *the load index*, *the geometry index* and *the rating index*.

Load index U_L relates to the gear size and to the load transmitted by the gear.

$$U_L = \frac{W_t \cdot P_{nd}}{F} \tag{2-8}$$

where W_t is tangential driven load, P_{nd} is the normal diametral pitch and F is contacting face width.

Geometry index C_t is related to gear teeth characteristics such as tooth size, number of teeth, root fillet radii, etc.

$$C_t = \frac{K_f \cdot m_N}{Y_c \cdot \cos\psi} \tag{2-9}$$

where K_f is a stress correction factor, m_N load sharing ratio, Y_c tooth form factor and ψ the helix angle.

Rating index R_A is related to the teeth misalignment, relative velocity of the mesh, manufacturing inaccuracies, etc.

$$R_A = \frac{K_o + K_m + K_s}{K_v} \tag{2-10}$$

where K_o is an overload factor, K_m alignment factor, K_s size factor and K_v velocity factor.

The *principal bending stress* or *beam bending stress* is given by the following equation:

$$S_t = U_L \cdot C_t \cdot R_A \tag{2-11}$$

2.7.3.2 Contact stresses. Hertzian theory

The shape and stresses of gear tooth contact are subject to the theory postulated by Heinrich Hertz in 1881. When two non-conforming solids are brought into contact they touch at a single point or along a line. Under the action of the load they deform in the vicinity of their point of first contact so that they touch over an area, which is very small compared to the dimensions of the two bodies. Hertz derived the deformation of the surfaces, strains and stresses for two bodies and showed that the shape of the contact area is elliptical and the contact pressure distribution is semi-ellipsoidal.

Hertz assumptions are as follows:

- Each body is approximated by an elastic half-space;
- The dimension of the contact area is small compared to the dimensions of the bodies and to the radii of curvature of the surfaces;
- The strains are sufficiently small so that linear elasticity is valid;
- There is no friction within the contact.

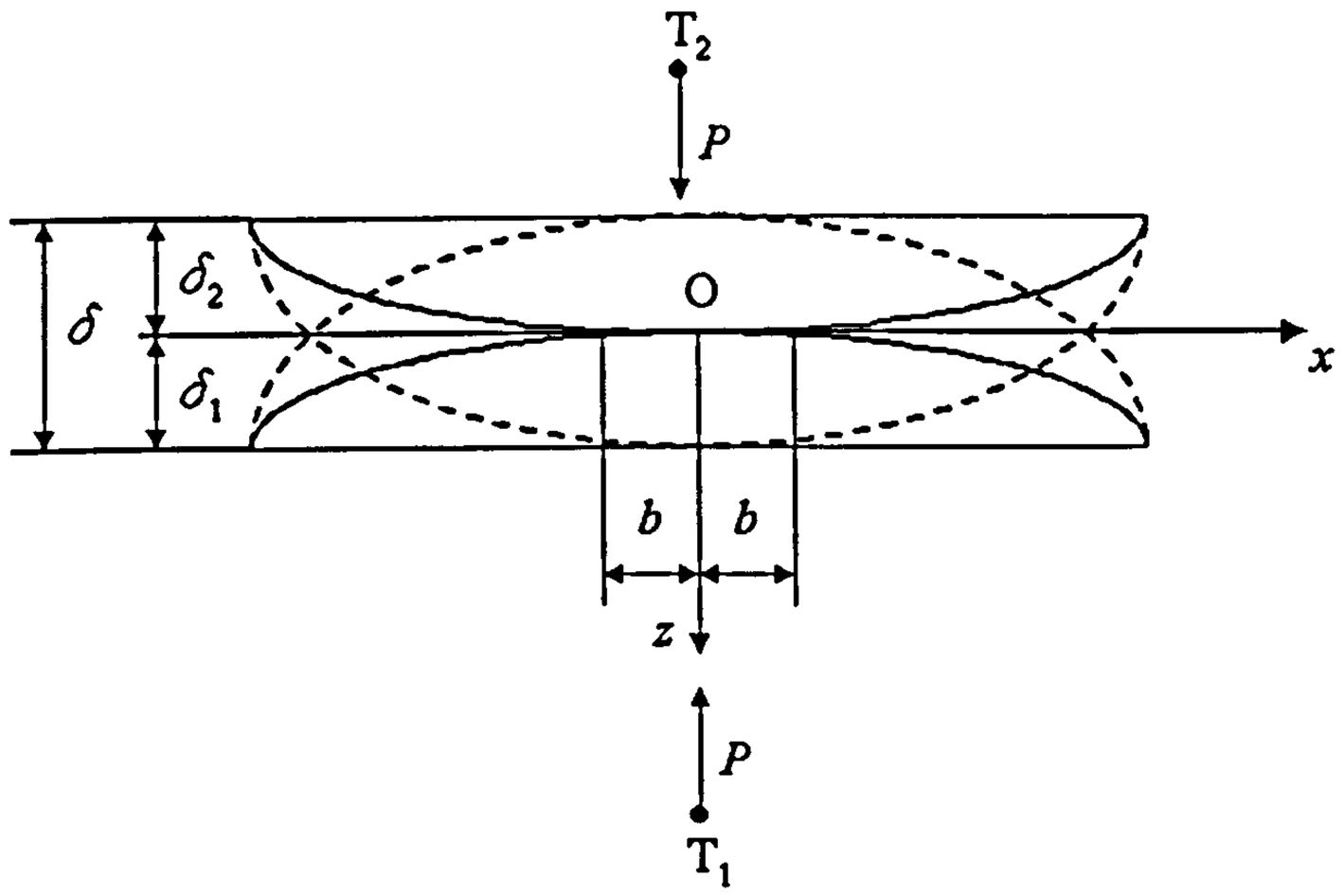


Figure 2-20. Hertzian contact between two non-conforming elastic bodies.

Figure 2-20 shows the deformation of the two solids under normal load *P*. Using Johnson's notations (Johnson, 1985), during compression the distant points T_1 and T_2 displace the distances δ_1 and δ_2 parallel to the *z*-axis towards O. The total displacement, $\delta = \delta_1 + \delta_2$ is called the *normal approach of interference.*

2.7.3.2.1 Circular contact

For the case of solids of revolution (e.g. spheres), the contact area is circular. The interference, δ, contact radius, a, and maximum contact pressure, p_0 (which occurs at $x = 0$) are given by the following equations:

$$\delta = \left(\frac{9P^2}{16R_e E'^2} \right)^{\frac{1}{3}} \tag{2-12}$$

$$a = \left(\frac{3PR_e}{4E'} \right)^{1/3} \tag{2-13}$$

$$p_0 = \left(\frac{6PE'^2}{\pi^3 R_e^2} \right)^{1/3} \tag{2-14}$$

where R_e is the effective (or composite) radius of curvature and is given by the equation (2-15) and E' is the elastic contact modulus (also known as composite or reduced Young's modulus) and is given by the equation (2-16).

$$\frac{1}{R_e} = \frac{1}{R_1} + \frac{1}{R_2} \tag{2-15}$$

$$\frac{1}{E'} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \tag{2-16}$$

where, R_1 and R_2 are the radii of curvature of the spheres, E_1 and E_2 their respective elastic moduli and ν_1 and ν_1 the Poisson's ratios of the two spheres.

2.7.3.2.2 Line contact

When two elastic cylinders are pressed together they flatten over a *rectangular area* of contact with a width of $2b$ as shown in Figure 2-21. The maximum contact pressure, p_0, and the semi-width of contact, b, are given by the Equations (2-17) and (2-18).

$$p_0 = \left(\frac{P' E'}{\pi R_e} \right)^{\frac{1}{2}} \tag{2-17}$$

$$b = \left(\frac{4P' R_e}{\pi E'} \right)^{\frac{1}{2}} \tag{2-18}$$

where P' is the load per unit length.

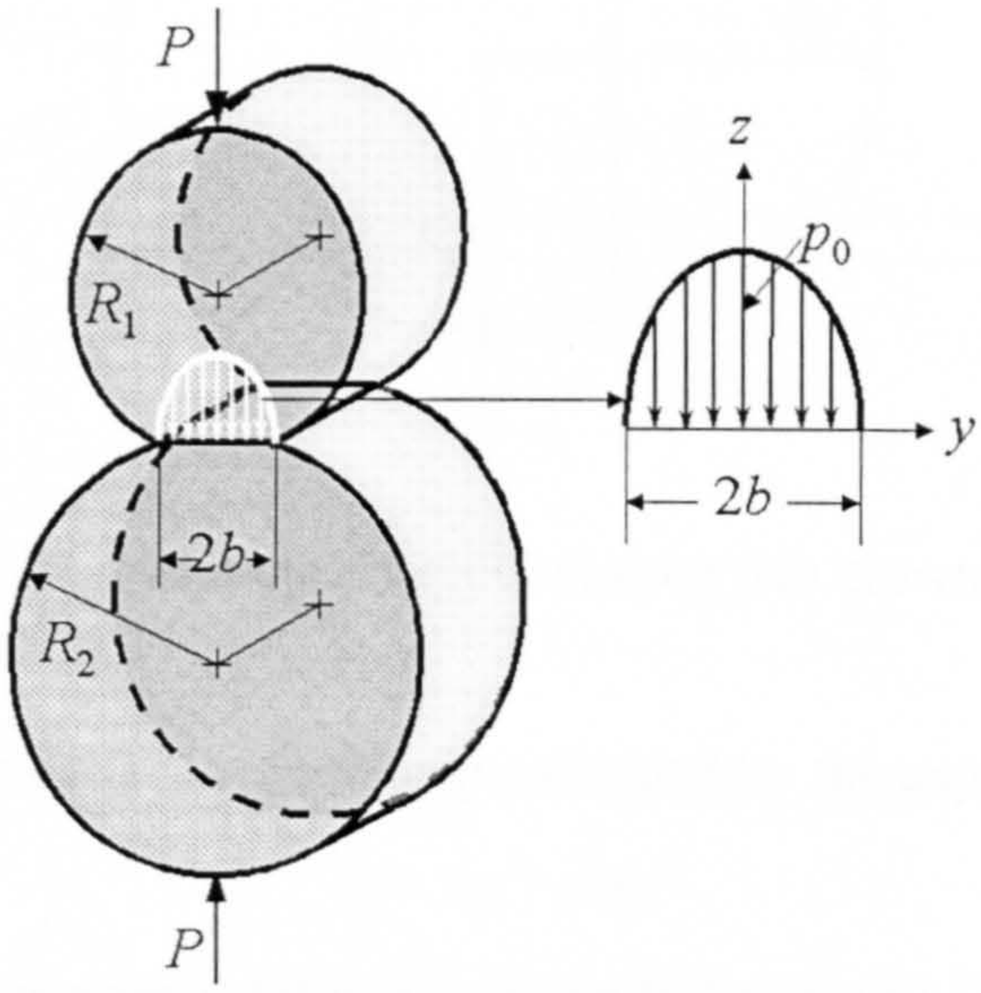


Figure 2-21. Hertzian contact of two cylinders. The area of the contact is rectangular and the pressure has an elliptical distribution. P is the applied load, p_0 the maximum contact pressure and $2b$ contact width.

The maximum shear stress and the depth of maximum shear stress are given by equations (2-19) and (2-20) respectively.

$$\tau_{max} = 0.30 p_0 \tag{2-19}$$

$$z = 0.78b \tag{2-20}$$

2.7.3.2.3 Elliptical contact

In practice, the cylinders (discs) used to simulate gear tooth contact are, almost always crowned due to manufacturing (e.g. grinding). The contact area of such cylinders is an ellipse with major semi-axis a and minor semi-axis b. In Figure 2-22, R_1' and R_2' are the maximum radii of curvature of cylinder 1 and 2, and R_1'' and R_2'' are the minimum radii of curvature of the respective cylinders.

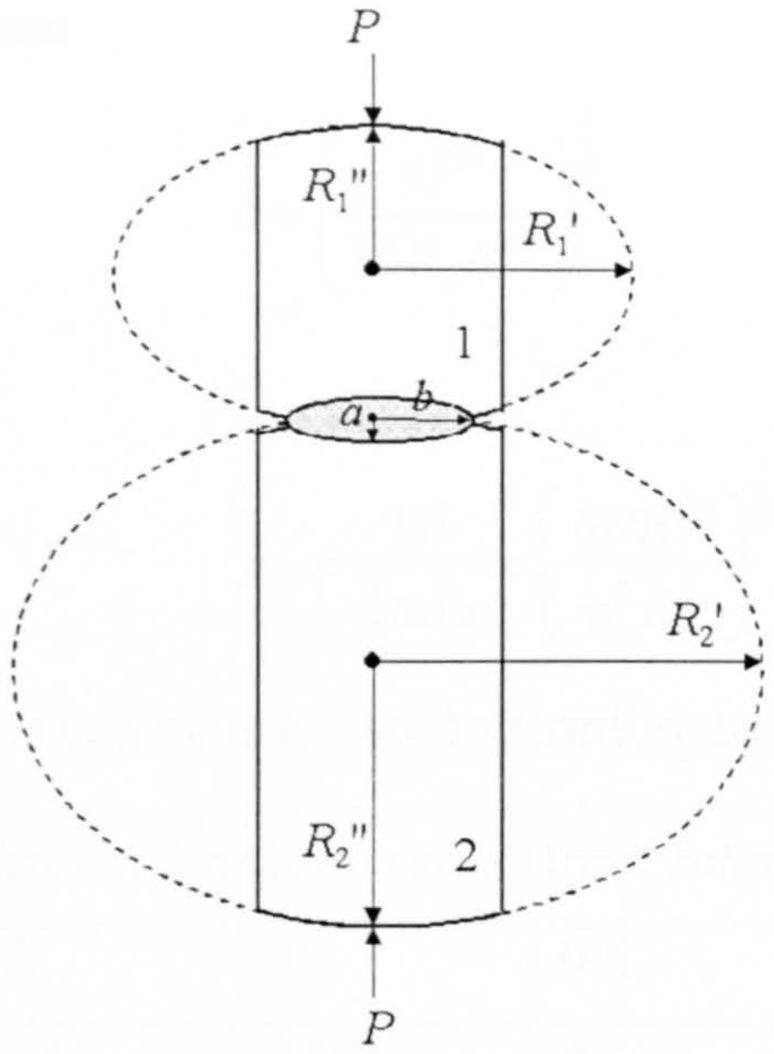


Figure 2-22. Elliptical contact between two crowned discs.

The separation between surfaces can be approximated by the expression below:

$$h = Ax^2 + By^2 \tag{2-21}$$

where A and B are two constants defined as in the following expressions:

$$A = \frac{1}{2R'} \text{ and } B = \frac{1}{2R''} \tag{2-22}$$

where R' and R'' are the *major* and *minor relative radii of curvature* given by

$$\frac{1}{R'} = \frac{1}{R_1'} + \frac{1}{R_2'} \tag{2-23}$$

$$\frac{1}{R''} = \frac{1}{R_1''} + \frac{1}{R_2''} \tag{2-24}$$

The ratio between the semi-axes of ellipse can be approximated as

$$\frac{a}{b} \approx \left(\frac{R'}{R''}\right)^{\frac{2}{3}} \tag{2-25}$$

The equivalent radius of curvature is defined by the relation:

$$R_e = \sqrt{R' \cdot R''} \tag{2-26}$$

and the ellipticity as

$$c = \sqrt{ab} = \left(\frac{3PR_e}{4E'}\right)^{\frac{1}{3}} \tag{2-27}$$

The approach of distant points

$$\delta = \left(\frac{9P^2}{16R_e E'^2} \right)^{\frac{1}{3}} \tag{2-28}$$

Maximum contact pressure:

$$p_0 = \frac{3P}{2\pi ab} = \left(\frac{6PE'^2}{\pi^3 R_e^2} \right)^{\frac{1}{3}} \tag{2-29}$$

Maximum shear stress values can be calculated according to Table 2-1.

Table 2-1. Maximum shear stress (from Johnson, 1985).

b/a	0	0.2	0.4	0.6	0.8	1.0
z/b	0.785	0.745	0.665	0.590	0.530	0.480
τ_{max}/p_0	0.300	0.322	0.325	0.323	0.317	0.310

In ordinary gears the maximum shear stress exists at about 0.18-0.3 mm below the surface and ahead of the contact point (Fernandes & McDuling, 1997).

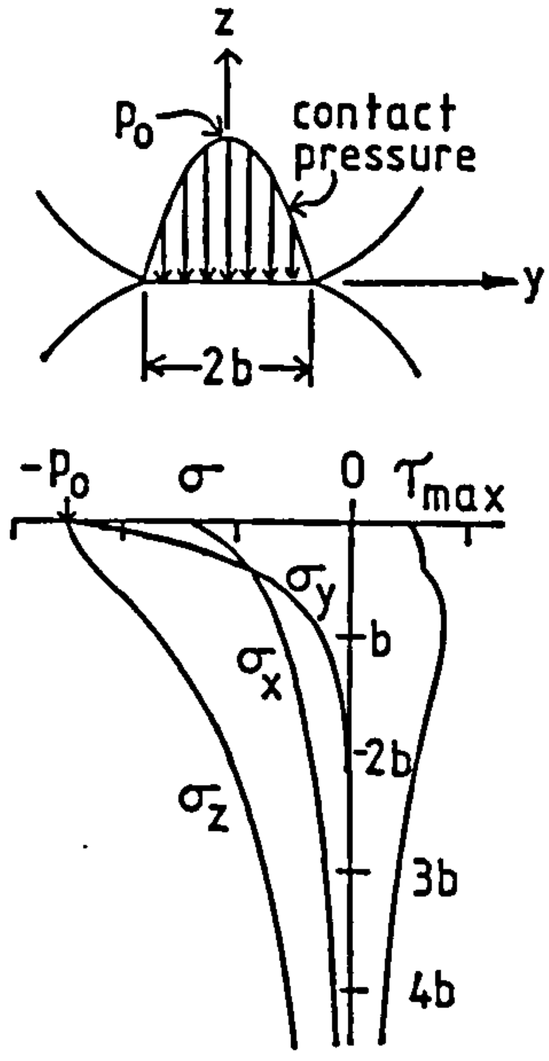


Figure 2-23. The variation of stresses as a function of depth below the contact area (from R. Priestner & D.M. Priestner, 1991).

A very useful investigation has been carried out by Dabnichki & Crocombe (1999) by comparing stresses predicted by different standards, which used Hertz theory, and stresses predicted by a finite element modelling (FEM) that included gear dynamics,

worked out by themselves. They found the values of the compressive stress provided by BS 436: 1986 are about 43 % below the values arising from their calculation.

Hertz theory applies to perfectly smooth surfaces. In practice, a solid surface is rough, irrespective of its method of formation. It contains valleys and hills (asperities), and the area of contact is reduced to a summation of asperity contacts. The concept of roughness is detailed in the next section and rough surface stress distribution is presented in Chapter 3. Besides the Hertzian stresses, the asperity contacts also generate stresses, which are located below but very close to the surface. Therefore, below the surface there are two zones where the shear stresses reach the maximum: one is due to the geometry of the surface (macrocontact or Hertzian contact) and the other is due to asperities contact (microcontact). It is generally believed that fatigue failure is initiated in one of these regions but micropitting occurs due to asperity contacts (Olver, 1983 & 1995). More details are given in Chapter 3.

2.8 Surface roughness

The boundary that separates a solid body from the surrounding medium (gas, liquid or solid) is called the *real surface* of the body. An *ideal surface* can be defined as a perfect smooth surface. In practice, when an engineering product is designed the intended surface, which is specified by the drawing and/or by other technical documents is the ideal surface and is referred as the *nominal surface* (BS 6741:1987). The real surface is deviated from the nominal surface as a result of the manufacturing process by which the surface has been created. Characterisation of the real surface can be done by breaking down its deviations from the nominal surface into measurable components. The measurable surface is known as *surface texture*. Surface texture is of crucial importance because it affects real area of contact, friction, wear and lubrication. Gear manufacturing processes leave relatively rough surfaces; typical high quality ground gears used in the aerospace industry have a surface roughness of about $R_a \sim 0.4$ μm (Britton et al., 2000).

2.8.1 Surface texture characterisation

It is postulated that the surface profile consists of a superposition of an infinite number of waves (Greenwood, 1992). Surface texture components are classified as a function of their wavelength as follows: *roughness*, *waviness*, *lay*, and *flaws* (Figure 2-24). A representation of the real surface (surface texture) that is made in two dimensions is called *surface profile* and in three dimensions a *surface map*.

Roughness (nano- and micro-roughness) comprises the short wavelenghts fluctuations of the surface height characterised by hills or asperities (local maxima) and valleys (local minima). Asperities are referred to as *peaks* in a profile and *summits* in a surface map.

Waviness (macro-roughness) is formed by longer wavelength deviations of the surface height.

Lay is the predominant direction of the surface pattern, normally determined by the production method used.

Flaws are accidental interruptions in the texture (scratches, burrs, etc.) and they are not considered to be part of the texture.

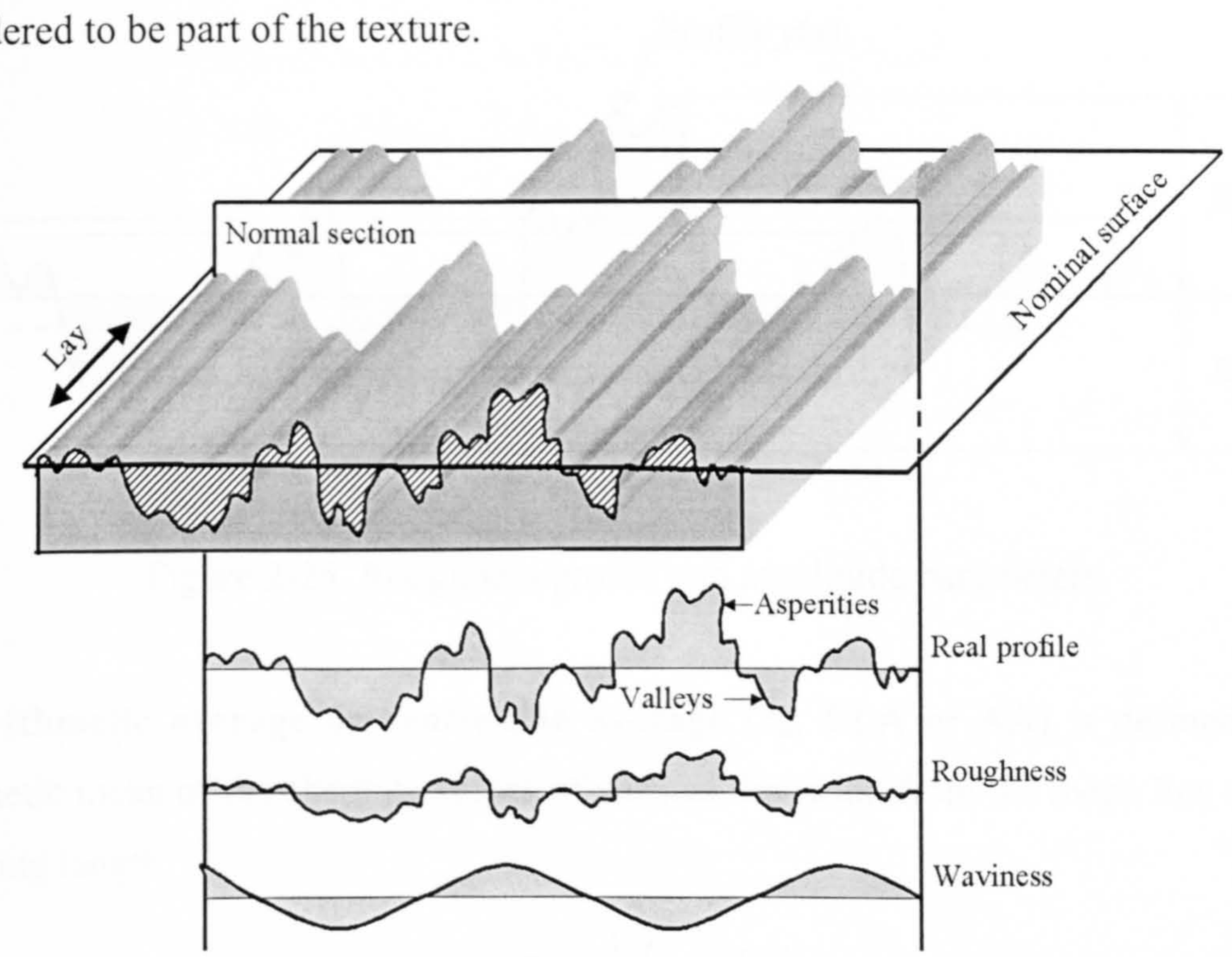


Figure 2-24. Surface texture characterisation.

Roughness parameters most commonly refer to the fluctuations in the surface height. Therefore, they are measured relative to a datum. The datum represents a line in a surface profile and a set of parallel lines in a surface map. The most common used methods to determine the reference line are: straight line fits, polynomial fits, low-pass filters and envelopes. Details can be found in Whitehouse, 1994. Most often the reference line called *the mean line* or *centre line* is determined by the least squares fitting method such that the area between the profile and the mean line above the line is equal to that below the mean line (see Figure 2-25).

2.8.1.1 Amplitude parameters

Currently, more than 30 parameters and functions of the surface have been developed. The most commonly used profile parameters, which give information about fluctuations in the surface height, are given below. Average roughness parameters for surface maps are calculated using the same mathematical approach.

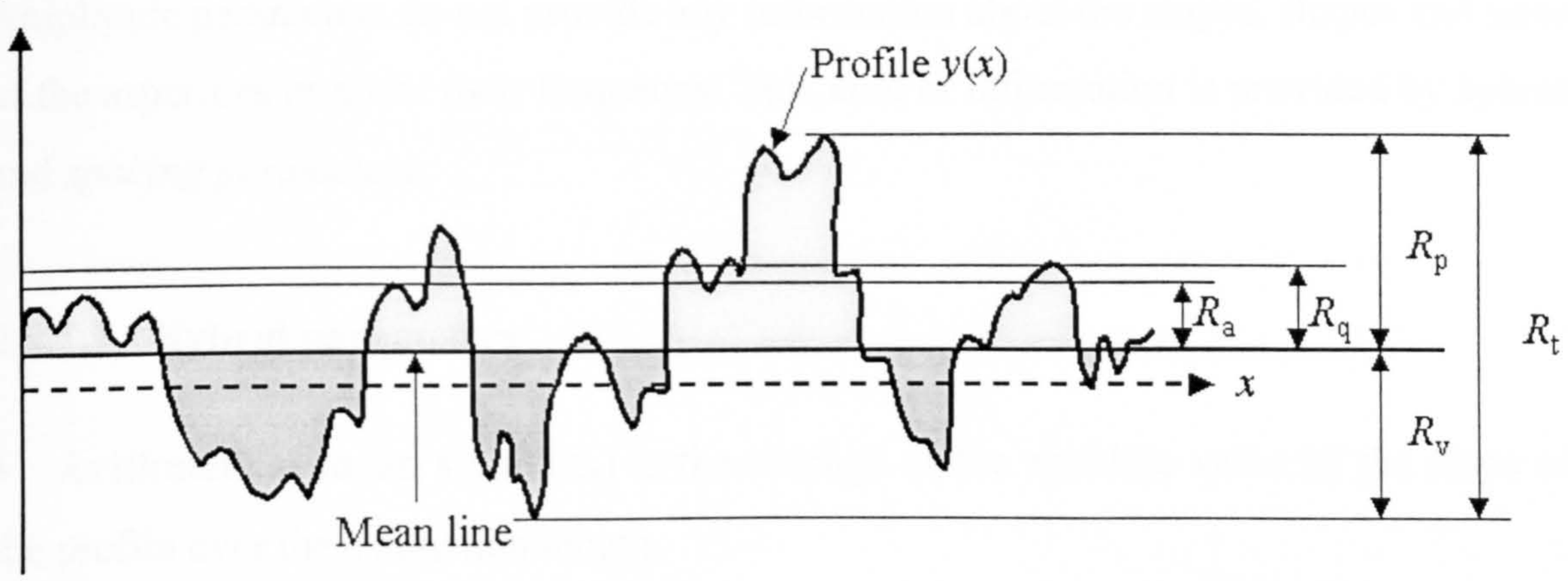


Figure 2-25. Roughness profile and amplitude parameters.

- **Arithmetic average or centre-line average** (R_a, CLA or AA) is defined as the arithmetic mean of the absolute values of vertical deviation from the mean line over the sampling length.

$$R_a = \frac{1}{l}\int_0^l |y(x)|dx \tag{2-30}$$

where $y(x)$ is a function that describe the surface profile, and l is the sampling length.

- **Standard deviation or variance or root mean square** (σ, R_q or RMS) is the square root of the arithmetic mean of the square of the vertical deviation from the mean line.

$$R_q = \sqrt{\frac{1}{l}\int_0^l y^2(x)dx} \tag{2-31}$$

- **Maximum peak height** (R_p) is defined as the distance between the highest asperity and the mean line.

$$R_p = \left|\max_{0<x<l}[y(x)]\right| \tag{2-32}$$

- **Maximum valley depth** (R_v) is defined as the distance between the mean line and the lowest valley.

$$R_v = \left|\min_{0<x<l}[y(x)]\right| \tag{2-33}$$

- **Peak-to-valley distance** (R_t, R_y, R_{max} or P-V) is the distance between the highest asperity and the lowest valley.

$$R_t = R_p + R_v \tag{2-34}$$

Amplitude parameters do not provide any information about the slopes, shapes and sizes of the asperities or about their frequency. This kind of information is provided by *hybrid* and *spacing parameters*.

2.8.1.2 Hybrid parameters

- **Arithmetical mean slope** (Δ_a) is the average of the absolute value of the slope of the profile over the evaluation length.

$$\Delta_a = \frac{1}{l}\int_0^l \left|\frac{dy}{dx}\right| dx \tag{2-35}$$

where, $\frac{dy}{dx}$ is the instantaneous slope of the profile.

- **Root-mean-square slope** (Δ_q or σ_m) is the root-mean-square value of the slopes within the sampling length.

$$\Delta_q = \sigma_m = \sqrt{\frac{1}{l}\int_0^l \left(\frac{dy}{dx}\right)^2 dx} \tag{2-36}$$

- **Mean radius of curvature**

$$\rho = \frac{1}{l}\int_0^l \left(\frac{d^2 y}{dx^2}\right) dx \tag{2-37}$$

- **Root-mean-square radius of curvature**

$$\sigma_k = \sqrt{\frac{1}{l}\int_1^l \left(\frac{d^2 y}{dx^2}\right)^2 dx} \tag{2-38}$$

2.8.1.3 Spacing parameters

- **Average wavelength (λ_a)**

$$\lambda_a = 2\pi\frac{R_a}{\Delta_a} \tag{2-39}$$

- **Root-mean-square wavelength (λ_q)**

$$\lambda_q = 2\pi\frac{R_q}{\Delta_q} \tag{2-40}$$

2.8.1.4 Amplitude distribution function

The amplitude probability distribution function (ADF) or **cumulative distribution function (CDF)** associated with the variable $y(x)$ is defined as the probability of the event $y(x) \le h$, and is written as:

$$P(h) = P(y \le h) \tag{2-41}$$

The function given by the derivative $p(y)$ is called the **probability density function (PDF)** and has the following expression:

$$p(y) = \frac{dP(y)}{dy} \tag{2-42}$$

There are two important parameters that characterise the PDF: skewness (*Sk*) and kurtosis (*ku*).

- **Skewness** (*Sk*) is a measure of asymmetry of the probability density function.

Gaussian distribution functions have zero skewness. Positive skewness indicates that the PDF bell is bent to the left (i.e. the surface has more valleys than peaks) and negative

skewness indicates the bell is bent to the right (i.e. the surface has more peaks than valleys).

$$Sk = \frac{1}{\sigma^3 l}\int_0^l y^3(x)\mathrm{d}x \tag{2-43}$$

- **Kurtosis** (*ku*) is a measure of the degree of pointedness or bluntness of the probability distribution function (PDF). It indicates the shape of the peak height distribution in terms of width and height. Gaussian profiles have a kurtosis of 3. PDF curves with kurtosis less than 3 are blunt and those with kurtosis higher than 3 are sharp.

$$ku = \frac{1}{\sigma^4 l}\int_0^l y^4(x)\mathrm{d}x \tag{2-44}$$

Gear tooth surfaces produced by grinding, are characterised by negative skewness and high kurtosis (see Figure 2-26) values which indicate a large number of asperities (negative skewness) lying in a narrow range of heights (kurtosis >3).

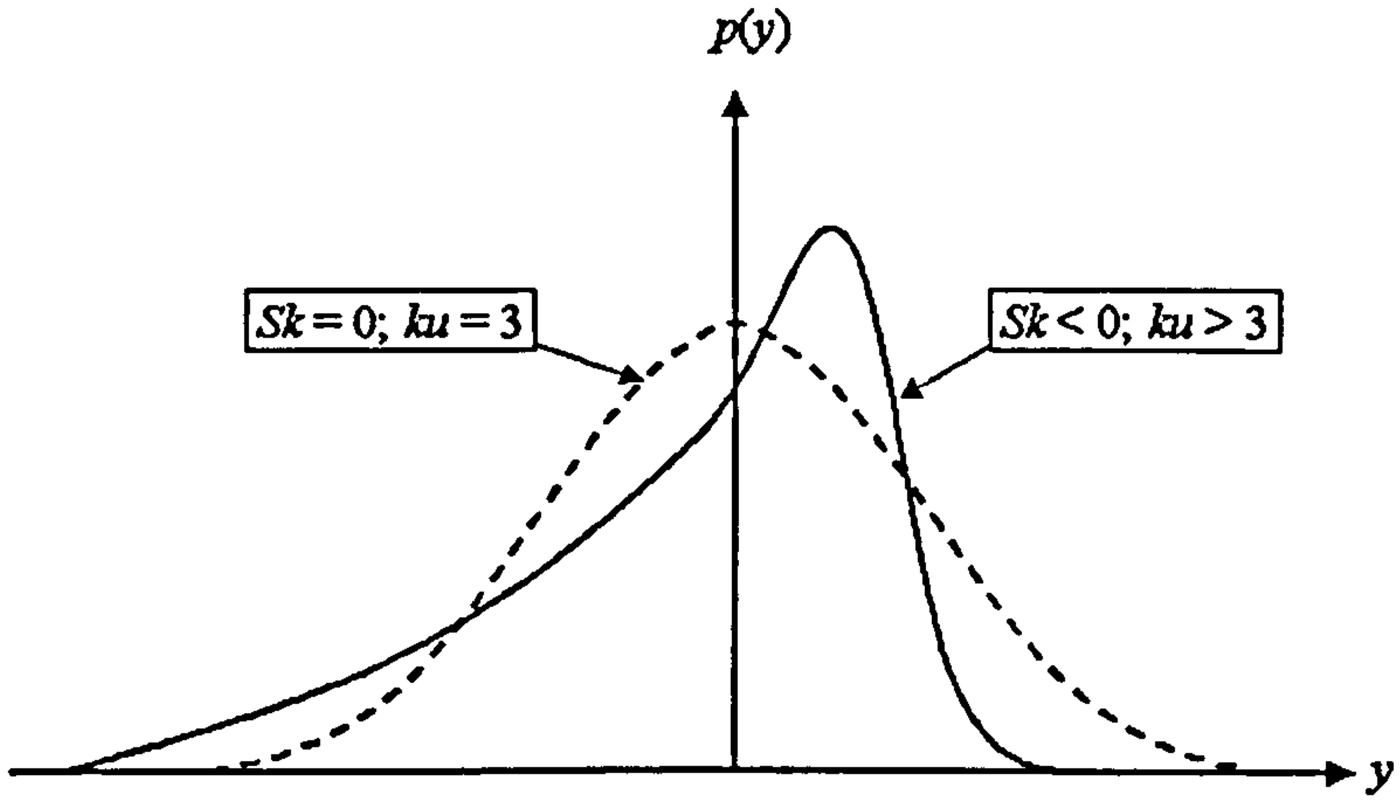


Figure 2-26. Probability density function (PDF) and its parameters: skewness and kurtosis. Dashed curve represents a Gaussian function. PDF for gear tooth surfaces (solid curve) have negative skewness (large number of asperities) and kurtosis higher than 3 (asperities height lie in a narrow range of values).

With the parameters described so far a surface profile (map) only consists of asperities and valleys. But are all the asperities significant? From the point of view of contact mechanics further "breakdown" of the profile is necessary in order to make distinction between peaks that deform plastically under load and the peaks that carry the load. This is done by the *bearing area curve* (BAC) also known as *Abbot-Firestone curve* or *material ratio curve.*

2.8.1.5 Bearing area curve (BAC)

As mentioned in the previous chapter, due to roughness the real area of contact between two gear teeth is smaller than that resulting from Hertzian calculations. A method to determine the amount of material in contact from a profile or a surface map has been first described by Abbot & Firestone (1933). A bearing area curve of a profile is produced by drawing parallel lines to the reference line. The lengths situated inside the profile are measured and summed. The proportion of the sum to the total length is calculated and plotted as a function of height. The resulting curve is the bearing area curve (Figure 2-27) and it has an S-shape for Gaussian surfaces.

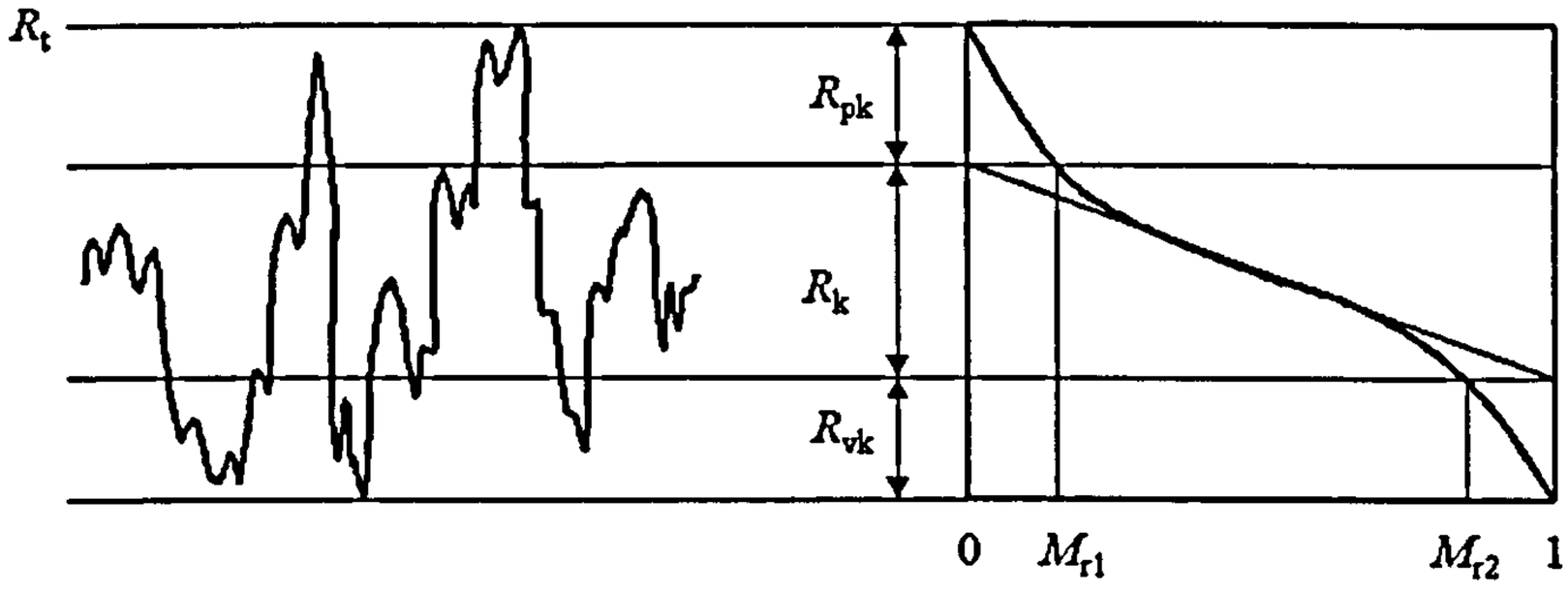


Figure 2-27. Bearing area curve (Abbot-Firestone curve or material ratio curve) and its parameters.

BAC parameters are determined by the minimum secant slope that has a length of 40 % of the peak-to-valley distance (P-V) when projected on the horizontal axis.

- **Core roughness depth** (R_k) is the depth of the roughness core profile.
- **Reduced peak height** (R_{pk}) is the mean height of the peaks protruding from the core area. It represents an estimate of the small peaks that will be worn off during the run-in period for a gear.
- **Reduced valley depth** (R_{vk}) is the mean depth of the valleys extending from the core area. It represents an estimate of the depth of valleys which will retain lubricant in a functioning gear.
- **Minimum material ratio** (M_{r1}) is the fraction of the surface which consists of small peaks above the core roughness.
- **Maximum material ratio** (M_{r2}) is the fraction of the surface which will carry load during running of gear. (1-M_{r2}) is the fraction of the surface that will retain lubricant.

- **"Area" of the peaks (A_1)**

$$A_1 = \frac{1}{2} R_{pk} \cdot M_{r1} \tag{2-45}$$

- **"Area" of the valleys (A_2) or oil retention "volume"**

$$A_2 = \frac{1}{2} R_{vk} \left(1 - M_{r2}\right) \tag{2-46}$$

This approach can only be used if the BAC is S-shaped. This is usually the case with ground surfaces such as gear tooth surfaces.

2.8.1.6 Statistics of asperities and valleys

In order to process the data about asperities statistically a more precise definition of the asperity concept is needed. In a profile, a *peak* is defined as a point higher than its two adjacent points greater than a threshold value. In a surface map, a *summit* is defined as a point higher than its four adjacent points greater than a threshold value. A valley is defined in the same way but in the reverse sense.

The statistics of asperities is taken to mean the assessment and representation of asperity distribution curves, which can be done in similar way to surface height, slope and curvature distribution.

2.8.1.6.1 Autocorrelation Function (ADF) and Power Spectral Density Function (PSDF)

A single profile represents a very small subset of the total surface texture data available. One way to increase the predictability of the measurements of the random surfaces is by using the autocorrelation function (ADF) and power spectral density function (PSDF). Ample explanations can be found in Whitehouse, 1994. The most important definitions are given below. If y_1 and y_2 are two measurements taken at a distance τ apart then the ADF is given by:

$$A(\tau) = E[y_1, y_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 p(y_1 y_2) dy_1 dy_2 \tag{2-47}$$

where, E denotes the expected value of the measurements taken over all the surface.

The PSDF is related to ADF by means of Fourier transforms and is given by:

$$P(\omega)=\int_{-\infty}^{\infty}A(\tau)\cos(\omega\tau)\mathrm{d}\tau \tag{2-48}$$

where ω is the angular frequency. Some features of tribological importance can be estimated using the **spectral moments of 0, 2nd and 4th order** of the PSDF: m_0, m_2, m_4 and the **bandwidth parameter** α:

$$m_0=\sigma^2=\frac{1}{l}\int_0^l y^2\mathrm{d}x=R_q^2 \tag{2-49}$$

$$m_2=(\sigma')^2=\frac{1}{l}\int_0^l\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2\mathrm{d}x=\Delta_q^2=\sigma_m^2 \tag{2-50}$$

$$m_4=(\sigma'')^2=\frac{1}{l}\int_0^l\left(\frac{\mathrm{d}^2y}{\mathrm{d}x^2}\right)^2\mathrm{d}x=\sigma_k^2 \tag{2-51}$$

$$\alpha=\frac{m_0\cdot m_4}{m_2^2}=\frac{\sigma^2\sigma_k^2}{\sigma_m^4} \tag{2-52}$$

where $\sigma = R_q$ is the root-mean square height, σ' is the standard deviation of the first derivative and represents the root-mean-square slope, Δ_q or σ_m and σ'' is the second derivative and represents the root-mean-square curvature, σ_k. These parameters are sufficient to characterise the asperities by the means of the profile parameters (Greenwood, 1992).

2.8.1.6.2 Asperity parameters

- **Peak density** is defined as the number of peaks of the profile per unit length.

$$D_p=\frac{1}{2\pi}\frac{\sigma_k}{\sigma_m} \tag{2-53}$$

- **Summit density** (μ_s) is the number of summits of surface map per unit area.

$$\mu_s=\frac{1}{6\pi\cdot\sqrt{3}}\left(\frac{\sigma_k}{\sigma_m}\right)^2 \tag{2-54}$$

- **Mean peak height**

$$r_p=\sqrt{\frac{\pi}{2\alpha}}=1.2533\frac{\sigma_m^2}{\sigma_k} \tag{2-55}$$

- **Mean summit height**

$$r_s = \frac{4\sigma}{\sqrt{\pi\alpha}} = 2.2568\frac{\sigma_m^2}{\sigma_k} \tag{2-56}$$

- **Standard deviation of peak heights**

$$\sigma_p = \left(1 - \frac{0.5708}{\alpha}\right)\sigma^2 \tag{2-57}$$

- **Standard deviation of summit heights**

$$\sigma_s = \left(1 - \frac{0.8468}{\alpha}\right)\sigma^2 \tag{2-58}$$

- **Mean peak curvature**

$$k_p = \sigma_k\sqrt{\frac{\pi}{2}} = 1.2533\sigma_k \tag{2-59}$$

- **Mean summit curvature**

$$\bar{k}_s = \beta^* = \frac{8\sigma_k}{3\sqrt{\pi}} = 1.5045\sigma_k \tag{2-60}$$

- **Standard deviation of peak curvature**

$$\sigma_{k_p} = 0.4242\sigma_k^2 \tag{2-61}$$

- **Standard deviation of summit curvature**

$$\sigma_{\beta^*} = 0.5178\sigma_k^2 \tag{2-62}$$

From the point of view of gear tooth contact the characterisation of individual surfaces must be extended to the characterisation of two rough surfaces brought into contact. To quantify their behaviour only three parameters must be known (Johnson, 1985): *the standard deviation of the asperity heights* σ_s, *the mean curvature of summits* $\bar{k}_s$, and *the asperity density* μ_s.

2.8.2 Contact between rough surfaces

The concept of asperity is as old as the first and second laws of friction. Amontons (1699) suggested that the friction between two bodies in sliding is caused by the force required to lift interlocking asperities over each other. When two surfaces are pressed together, the first contact takes place at the top points of the highest asperities, which

deform plastically or elastically under the load so more asperities will be brought into contact. Therefore the real contact area may be only about 10 % of the surface area predicted by Hertzian theory (Holmberg & Matthews, 1994). Depending on the approach used to quantify the contact between rough surfaces two types of models have been described in literature: *single asperity* and *multi-asperity models*. They do not exclude each other but they are complementary. Equations derived for a single pair of asperities are then applied to a large number of contacts. If no interaction between asperity pairs is considered then the multi-asperity model is called *uncoupled* and when the equations of elasticity are solved simultaneously for the entire body the multi-asperity model is called *coupled*. A review of all asperity models is given in Adams & Nosonovsky (2000).

Uncoupled multi-asperity models predict that the real area of contact, A_c, is proportional to $P^{2/3}$, where P is the applied load if the number of asperities in contact is constant and they deform elastically. If the number of asperity contacts increase then A_c is proportional to P, regardless the deformation mode (elastic or plastic).

Coupled multi-asperity models solve the elasticity equations for the whole body by considering a periodic interface profile (e.g. sinusoidal) and add the following boundary conditions: normal displacements are continuous in the contact zone and normal stress is zero in the separation zone. The very first uncoupled model was suggested by Grenwood & Williamson in 1966.

2.8.2.1 Greenwood-Williamson asperity model

The Greenwood-Williamson (1966) model is extensively used in tribology because of its relative simplicity and its applicability to real surfaces. It is based on a geometrical model suggested in 1737 by Belidor (Dowson, 1979) according to which rough surfaces can be represented by spherical asperities. The model describes the contact between one rough surface and one smooth surface and is based on the following assumptions:

- The rough surface is isotropic;
- All asperities are spheres of the same radius of curvature R while their heights vary randomly;
- Summit heights have a Gaussian distribution;

- The load supported by asperities is not influenced by neighbouring asperities;

If $\Phi(z)$ is the height probability function and d is the separation distance between the reference plane of the rough surface and the rigid plane then all asperities with a height greater than d will make contact and will be compressed a distance $\delta = z - d$.

The area of individual contact is $A_i = \pi \cdot a_i^2$ where

$a_i = \left(\frac{3W_i R}{4E'}\right)^{\frac{1}{3}}$ is the radius of the contact circle, W_i is the load carried by asperity $i = 1...n$ and E' the reduced elastic modulus.

- **The number of contacts**

$$n = \int_d^\infty N\Phi(z)dz \tag{2-63}$$

- **Total area of contact**

$$A = \int_d^\infty N\pi R(z-d)\Phi(z)dz \tag{2-64}$$

where N is the total number of asperities.

- **Total load**

$$W = \int N \cdot \frac{4}{3} E' R^{\frac{1}{2}} (z-d)^{\frac{3}{2}} \Phi(z)dz \tag{2-65}$$

- **Plasticity index**

The mode of deformation of the asperities is analysed as a function of plasticity index Ψ, a factor that combines the material and surface roughness properties.

$$\Psi = \frac{E'}{H}\sqrt{\frac{\sigma_s}{R}} \tag{2-66}$$

where σ_s is the standard deviation of the summits height and H is hardness.

When $\psi < 1$ the asperity deformation is entirely *elastic*; when $\psi \geq 1$ the deformation is predominantly *plastic*. During the initial load of engineering surfaces the deformation will be predominately plastic (Adams & Nosonovsky, 2000).

Other formulas have been proposed for the plasticity index:

(McCool, 1986) $$\Psi = \frac{E^*}{H}\sqrt{\sigma_s \cdot k_s} \tag{2-67}$$

(Mikic, 1974)
$$\Psi = \frac{E' \cdot \sigma_m}{H} \tag{2-68}$$

Other multi-asperity models have been suggested or new considerations have been added to the existing models but the Greenwood-Williamson formulation remains one of the most convenient and accurate approaches.

2.8.2.2 Stresses below contacting asperities

It is well known that the region where failure is most likely to occur coincides with the depth of maximum Hertzian shear stress. It is also widely believed that plastic deformation of asperities followed by fracture and stresses generated at asperity contact play a key role in surface fatigue failure (Polonsky, 1995) particularly in micropitting (Ioannides & Kuijpers, 1986). Hertzian theory predicts stresses below the perfect smooth surfaces according to the equations given in section 2.7.3.2. A similar mathematical approach can be applied to calculate stresses under each local contact of any two asperities. Lo (1969) found that the mean contact pressure for two rough cylinders in contact is very large and nearly independent of the applied load (W) meanwhile for smooth cylinders the mean contact pressure is proportional to $W^{0.5}$ as the Hertz formulae show. The real area of contact and the number of contact spots increase almost linearly with the applied load. It follows that the contact area predicted by the rough surface model is much smaller (76 times in Lo's numerical example) than that predicted by smooth surfaces model (Hertz). Hence, the mean contact stress is very large (on the order of the hardness) for rough surfaces which increases the probability of plastic deformation of the asperities.

Students & Rudzitis (1996) have shown that the most critical zones may be both, within the surface layer and in the contact area itself. They determined the maximum stress below the surface, depending on the geometric parameters, the friction coefficient and the maximum stress in the centre of the contact area. The *maximum reduced equivalent stress* is given by:

$$\sigma_{e_{max}} = \left(0.251 + 0.016\frac{b}{a} + 1.158\mu^3\right) q_{max} \tag{2-69}$$

where a and b are the semi-axes of the contact area, μ is the friction coefficient and q_{max} is the maximum pressure in the centre of contact. As the friction coefficient increases the region moves toward the contacting surface.

2.9 Gear fluid lubrication

Steel or other metallic gears operating dry will wear out quickly due to high friction and high temperature generated by friction. In order to minimise surface damage by reducing friction and removing the heat generated within the contact, gears are lubricated. A lubricant represents any substance (solid or fluid) interposed between two surfaces in relative motion having the basic functions of friction reduction, heat removal and suspension of contaminants.

2.9.1 Friction

The amount of friction developed between two contacting surfaces depends on the type of frictional system, the pressure exerted between the contacting bodies, the nature of the surfaces and the rubbing conditions (Gunther, 1972). The degree to which friction develops is indicated by the amount of heat generated, the rate of surface wear, the power required to initiate and maintain motion, and the ratio of power output and power input. Energy can be saved by several tribological measures reducing losses caused by friction and wear. Although vehicle gear efficiency is very high (97 % according to Bartz, 1993) fuel consumption of a car can be reduced by lubricant related measures of up to 5.1 % (Bartz, 2000).

As discussed in the previous chapters the motion within the gear tooth contact is a combination of sliding and rolling. Therefore, there are two basic kinds of friction: *sliding* and *rolling friction*. Friction reaches its highest magnitude under dry sliding conditions. That is, on one hand because of the bonding force of molecular attraction and on the other hand because of the surface roughness. Molecular attraction increases with pressure and surface smoothness because the area of contact becomes proportionately greater. When sliding over each other the asperities of one surface will shear their opposing asperities. If the pressure is high enough local temperatures will

reach a level that causes plastic flow and local welding at asperity level may occur which bonds the two surfaces together. This may cause the scuffing failure.

Theoretically, under pure rolling conditions, friction would not exist, since a point or a line has no area. However, as previously mentioned, under the action of applied loads the initial point or line contact becomes elliptical or rectangular due to elastic deformation. The material displaced by deformation forms a temporary mound at the contact edge (Gunther, 1971). Rolling friction, then, is the sum of the resistances opposed by the mound and the surface asperities.

Frictional energy, according to the First Law of Thermodynamics is ultimately converted to thermal energy. Friction, as a force that resists motion is measurable by the *friction coefficient*, which has the mathematical expression as in equation 2-70.

$$\mu = \frac{F}{N} \tag{2-70}$$

Under dry conditions, friction coefficient may have very high values: 0.15-0.40 for sliding friction compared to 0.002-0.004 for rolling friction (Gunther, 1971).

In order to reduce friction and the temperature in a gear tooth contact, gears are lubricated. The friction coefficient under lubricated conditions depends on the lubrication mechanism achieved.

2.9.2 Lubrication mechanisms

The mechanisms of lubrication can be divided into two main groups (Holmberg & Matthews, 1994): *fluid pressure* or *thin film lubrication* and *surface film* or *thick film lubrication*.

Fluid pressure lubrication comprises those mechanisms in which the surfaces are kept apart from each other by fluid pressure created in the lubricant within the contact. The mechanisms involved are *hydrostatic*, *hydrodynamic* and *elastohydrodynamic lubrication*.

Surface film lubrication is defined as the group of lubrication mechanisms in which surfaces are partly kept apart from each other by a protecting film attached by chemical or physical bonds to the surfaces. This category includes *boundary* and *solid lubrication*.

In the present work only lubrication by fluids is involved. The four mechanisms that are commonly encountered in fluid lubrication of gears are described below. Different regimes with their corresponding values for lubricant *film thickness*, *h*, *specific film thickness*, λ, and *friction coefficient*, μ, are presented in Table 2-2.

Table 2-2. Lubrication regimes and lubricant film properties (after Holmberg & Matthews, 1994).

Lubrication regime	**Lubricant film properties**		
	Film thickness *h* (μm)	Specific film thickness (λ)	Friction coefficient (μ)
Boundary	0.005-0.1	<1	0.03-0.2
Mixed	0.01-1.0	1-4	0.02-0.15
Elastohydrodynamic	0.01-10.0	3-10	0.01-0.1
Hydrodynamic	1-100	10-100	0.001-0.01

2.9.2.1 Hydrodynamic lubrication

In hydrodynamic lubrication the pressure created within the lubricant is high enough to keep the solid surfaces entirely separated. The following requirements must be met for hydrodynamic lubrication to take place (Holmber & Matthews, 1994): the surfaces must be in relative motion, they must be converging in the direction of motion and the lubricant between must be viscous. The property governing the hydrodynamic lubrication is the *viscosity* of the fluid. The hydrodynamic pressure is low compared to the strength of the solids so local deformation of the solids is not significant. The hydrodynamic theory described by Reynolds in 1886 assumes that friction occurs only within the fluid film, and is a function of fluid viscosity. In hydrodynamic lubrication higher speed gives better lubrication while very low speed may cause lubrication failure (Lansdown, 1996).

2.9.2.2 Elastohydrodynamic lubrication

Contact pressures of about 1-2 GPa are commonly encountered in carburised gears. Under these conditions the viscosity of lubricant is increased so much (viscosity increases almost exponentially with pressure) that it behaves as a "solid" and the two

surfaces deform elastically due to lubricant pressure. This lubrication regime is called *elastohydrodynamic* (EHD or EHL). In the contact zone the lubricant pressure causes a further increase in viscosity. The high viscosity combined with the short time required to carry the lubricant through the contact area results in the impossibility of the lubricant to escape, and the surfaces will remain separated. Load has a small effect on film thickness because, at the pressures involved, the lubricant film is actually as rigid as the metal surfaces. Therefore, the main effect of a load increase is to deform the metal surfaces and increase the contact area, rather than decrease the film thickness.

2.9.2.3 Boundary lubrication

Under certain conditions (such as: shock loading, high load, low speed, and low viscosity) an intermittent contact between the surfaces occurs. Under these circumstances, the fluid film is no longer capable of adequately protecting the surfaces and the surfaces move so close to each other that asperity interaction becomes significant. This regime is called *boundary lubrication*. Very thin lubricant films are adsorbed to the solid surfaces and significantly inhibit asperity welding. Boundary lubrication is dominated by surface interaction phenomena between the thin layers of lubricants and the solids. The fundamental interaction mechanisms between thin lubricant layers and surfaces are *physical adsorbtion*, *chemical adsorbtion* and *chemical reaction*. Physical adsorbed molecules have week bonds (Van der Waals) and the film provides lubrication only at low loads and low sliding speeds. Chemical adsorbtion occurs when the molecules are held to the surface by chemical bonds. Films formed in this way provide good lubrication at moderate loads, temperatures and sliding velocities. Chemical reactions result in the formation of a chemical compound, which usually contains sulphur, chlorine or phosphorus atoms. The film formation is strongly activated at temperatures above 150°C (Holmberg & Matthews, 1994). The films formed by chemical reactions provide good lubrication for high loads, high temperatures and high sliding speeds.

2.9.2.4 Mixed lubrication

Mixed lubrication, also referred to, as *partial* or *micro-elastohydrodynamic lubrication* is a regime governed by a mixture of elastohydrodynamic and boundary lubrication. The surfaces are separated by a thin lubricant film but asperity interaction still takes place. In such contacts the roughness significantly affects the distribution of pressure, and pressures much higher than predicted for smooth surfaces are obtained even under conditions where a full fluid film is present. Because of its complexity the mixed lubrication mechanism is poorly understood although important progress has been made in the last decades (Kaneta & Nishikawa, 1999; Snidle et al, 1997; Chang & Zhao, 2000; Evans & Snidle, 1996; Chang, 1995; Guangteng et al, 2000).

2.9.3 Viscosity

Viscosity is by far the most important parameter for establishing the thickness, pressure, and temperature of a lubricant film in gear lubrication. Viscosity, as opposed to fluidity can be defined as a measure of the internal friction of a fluid. Internal friction occurs due to the movement of a layer of fluid in relation to another layer. Isaac Newton has defined the mathematical expression of viscosity (equation 2-71) by considering the model represented in Figure 2-28.

$$\frac{F}{A} = \eta \frac{\mathrm{d}v}{\mathrm{d}x} \tag{2-71}$$

where, A is the area of the fluid layer, F the force required to maintain the speed difference dv between the two layers, dx is the distance between the parallel planes and, η is a constant for a given material and is called its *viscosity*.

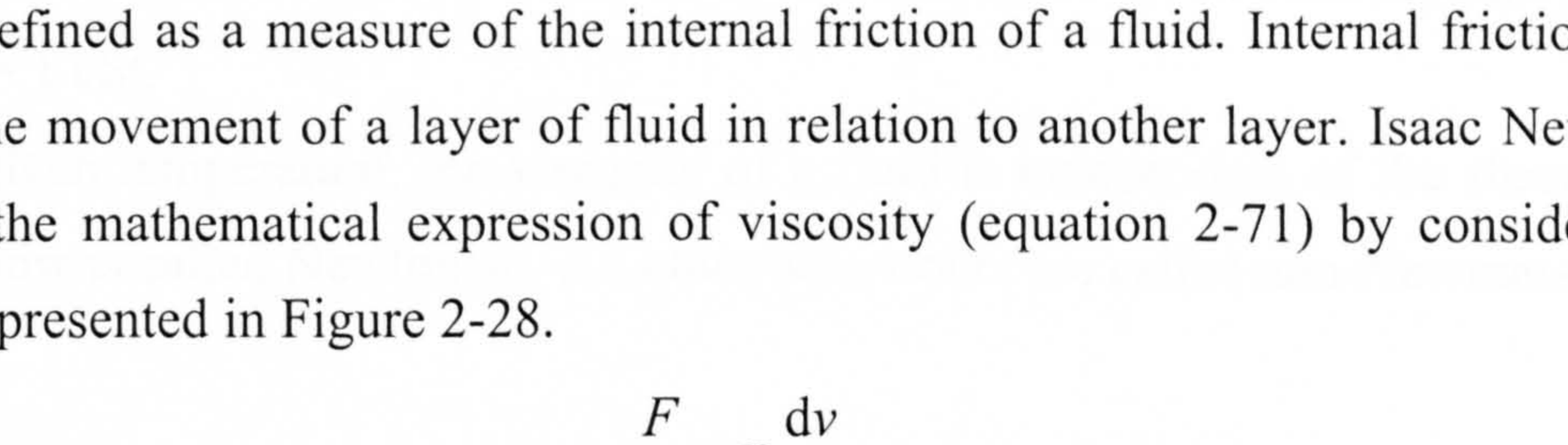


Figure 2-28. The concept of viscosity. Two parallel planes of fluid of equal area A are separated by a distance dx and are moving in the same direction at different velocities v_1 and v_2 (after Brookfield Eng, Inc., 2002).

The velocity gradient, $dv/dx = \gamma$, is a measure of the change in speed at which the intermediate layers move with respect to each other and is called the *shear rate*. The term $F/A = \tau$ indicates the force per unit area required to produce the shearing and is referred to as *shear stress*. Using these terms, viscosity can be defined mathematically as the ratio of shear stress to shear rate:

$$\eta = \frac{\tau}{\gamma} \tag{2-72}$$

Viscosity defined above is called dynamic or absolute viscosity and is usually reported in poise (P), or in SI units as pascal-seconds (Pa·s) where 1 Pa·s = 10 P. Because it is more convenient to report viscosity in a manner which accounts for the lubricant density, kinematic viscosity is normally used to characterise lubricants. Kinematic viscosity of a fluid, ν, equals its dynamic viscosity, η, divided by its density, ρ, at a given temperature.

$$\nu = \frac{\eta}{\rho} \tag{2-73}$$

Kinematic viscosity is usually reported in stokes (St), or in SI units as mm^2/s, where $1 mm^2/s = 1$ cSt.

If, at a given temperature, the viscosity of a fluid is independent of the shear rate the type of flow is called Newtonian. All other behaviours are called non-Newtonian.

2.9.4 Lubricant rheology

Rheology is the science involving the flow behaviour of substances. It deals with shear rate, shear stresses, and yield stresses of fluids. Lubricant rheology is taken to mean the relationship between the shear stress and the shear rate at different pressures and temperatures (Larsson et al., 2000).

At a given temperature, the viscosity of fluids such as water or thin motor oils is independent of the shear rate. This behaviour is called *Newtonian*. A *non-Newtonian* fluid is defined as one for which the ratio shear stress/shear rate is not constant. There are several types of non-Newtonian flow behaviour, characterised by the way a fluid's viscosity changes in response to variations in shear rate. The most common types of

non-Newtonian fluids include: pseudoplastic, dilatant, plastic, thixotropic and rheopexic. *A pseudoplastic* fluid will display a decreasing viscosity with an increasing shear rate. Increasing viscosity with an increase in shear rate characterises a *dilatant fluid*. The *plastic behaviour* implies that the fluid behaves as a solid under static conditions. The minimum force that must be applied to the fluid before any flow is induced is called the yield value. Some fluids will display a change in viscosity with time under conditions of constant shear rate. A thixotropic fluid undergoes a decrease in viscosity with time, while it is subjected to constant shearing. Rheopexy is the opposite of thixotropic behaviour, in that the fluid's viscosity increases with time as it is sheared at a constant rate.

The relationship between shear stress and shear rate in a gear lubricant can be expressed by the following equation (Mihailidis et al, 1999):

$$\gamma = A\frac{\mathrm{d}\tau}{\mathrm{d}t} + F(\tau) \tag{2-74}$$

where the first term of the sum describes the elastic behaviour, while the second the viscous behaviour. The most commonly used expressions of the second term in applications to gear lubrication are given below.

- Newton

$$F(\tau) = \frac{\tau}{\eta} \tag{2-75}$$

- Ree Eyring

$$F(\tau) = \frac{\tau_E}{\eta}\sinh\left(\frac{\tau}{\tau_E}\right) \tag{2-76}$$

- Bair Winer

$$F(\tau) = -\frac{\dot{\tau}_l}{\eta}\ln\left(1-\frac{\tau}{\tau_l}\right) \tag{2-77}$$

- Lee Hamrock

$$F(\tau) = \frac{\tau}{\eta\sqrt{1-(\tau/\tau_l)^2}} \tag{2-78}$$

where, τ_E and τ_l are limiting shear stresses.

2.9.5 The EHL contact equations

In order to solve the problem of an elastohydrodynamic contact of rough surfaces is necessary to take into account a certain number of phenomena. They are as follows (Ville, 1998): pressure generation within the contact, elastic deformation of the surfaces under contact pressure and the dependence of lubricant viscosity and density on pressure. By expressing these phenomena in a mathematical form it is possible to determine the lubricant film thickness and the pressure distribution within the contact. An elastohydrodynamic contact is fully described by the solutions of several equations simultaneously solved: the Reynolds equation, an equation that describes the piezoviscous behaviour of the lubricant (the dependence of viscosity on pressure, Roelands or Barus equation), an equation that describes the dependence of lubricant density on pressure (Dowson-Higginson equation), and an equation that describes the elastic deformations of the surfaces (Dowson-Higginson equation) (Ville, 1988).

2.9.5.1 The Reynolds equation

The equation which governs the pressure generation in lubricant is the Reynolds equation (Reynolds, 1886). The equation is derived by applying the basic equations of motion and continuity to the lubricant. It can be written in the following form:

$$\frac{\partial}{\partial x}\left(\frac{\rho}{12\eta}h^3\frac{\partial p}{\partial x}\right)=\frac{u_1+u_2}{2}\frac{\partial}{\partial x}(\rho h)+\frac{\partial}{\partial t}(\rho h) \qquad (2\text{-}79)$$

where: h lubricant film thickness, p pressure within lubricant, ρ the fluid density, η the dynamic viscosity of the fluid, x and y coordinates, t time, u_1, u_2 velocities of the two surfaces. Reynolds assumptions are as follows: the flow is laminar, the lubricant is Newtonian, the mass forces are negligible, the film thickness is very thin compared to the other contact dimensions, the deformations induced by the pressure are sufficiently small and they do not affect the speed.

2.9.5.2 The effect of pressure on viscosity

Theory assumes that the lubricant behaves as a Newtonian liquid with a viscosity that approximately follows the Barus equation (Barus, 1893) as an exponential variation:

$$\eta = \eta_0 e^{\alpha p} \tag{2-80}$$

where: η dynamic viscosity, η_0 dynamic viscosity at ambient pressure, α pressure-viscosity coefficient, p pressure.

For contact pressures higher than 1 GPa the pressure-viscosity dependence is better described by Roelands equation (Roelands, 1966).

$$\eta = \eta_0 \left(\frac{\mu_\infty}{\eta_0} \right)^{\left[1-\left(1+\frac{p}{c}\right)^2\right]} \tag{2-81}$$

where η_0 is reference viscosity, c a constant which depends on pressure.

A widely used relationship for pressure-viscosity coefficient is that of Wooster (Cameron, 1981).

$$\alpha = (0.6 + 0.965 \log \eta_0) \cdot 10^{-8} \ (\mathrm{Pa}^{-1}) \tag{2-82}$$

where η_0 is the viscosity at atmospheric pressure in mPas.

2.9.5.3 The effect of pressure on density

The dependence of lubricant density on pressure is given by the Dowson-Higginson (1977) equation:

$$\rho = \rho_0 \left[1 + \frac{0.6p}{1+1.7p} \right] \tag{2-83}$$

where ρ_0 is the density at ambient pressure and p pressure.

2.9.5.4 Elastic deformation of surfaces

Considering the following assumptions, deformations are entirely elastic so they obey Hooke's law, the materials are homogeneous and isotropic, the boundary conditions are of a solid half-space then the elastic equation can be written as below (Dowson & Higginson, 1977).

$$h = h_0 + \frac{x^2}{2R_e} - \frac{2}{\pi E'} \int_{s1}^{s2} p(s) \ln(x-s)^2 \, ds + \varepsilon \tag{2-84}$$

where h is the lubricant film thickness, h_0 film thickness on line of centres, R is the equivalent radius of curvature, E' is the reduced elastic modulus, x the direction of rolling, p pressure, s coordinate on x direction, and ε is a constant.

2.9.6 Lubricant film thickness

The first attempt to calculate the lubricant film thickness was that of Martin (1916) who solved the Reynolds equation for gear tooth contact conditions, with the assumption that the lubrication fluid is isoviscous. Martin's analysis enabled him to establish a formula for the predicted film thickness within the contact as:

$$h_{\min} = h_{cen} = 4.9UW_L^{-1} \tag{2-85}$$

where: $h_{\min}$ is the dimensionless minimum film thickness, h_{cen} is the dimensionless central film thickness, and U is the dimensionless speed parameter given by equation 2-86.

$$U = \frac{\eta_0 u}{E' R_e} \tag{2-86}$$

W_L is the dimensionless load parameter given by the following equation:

$$W_L = \frac{w}{LE' R_e} \tag{2-87}$$

E' is the elastic contact modulus, η_0 is the viscosity at ambient pressure, u the entraining velocity

$$u = \frac{u_1 + u_2}{2} \tag{2-88}$$

u_1, u_2 are the velocities of the two surfaces, R the equivalent radius of curvature, w the load, L the width of the line contact.

The film thickness predicted by Martin's equation was about 0.01μm, which is much smaller than the surface roughness of ordinary gears (about 0.4-0.8 μm), therefore his theory failed to explain the lubrication of gear teeth.

A major advance in the development of lubrication theory was made by Grubin (1949), who combined the elastic contact theory of Hertz with a lubrication analysis that included the effect of pressure on viscosity:

$$h_{cen} = 1.95U^{8/11}G^{8/11}W_L^{-1/11} \tag{2-89}$$

where: G is a dimensionless materials parameter and α is the pressure-viscosity coefficient

$$G = \alpha E' \tag{2-90}$$

Predicted film thicknesses calculated by Grubin were two orders of magnitude greater than those predicted by Martin's theory, comparable to the surface roughness of gears.

Petrusevich (1951) found a very important feature, that is near to the outlet end of the conjunction a local restriction in film thickness occurs and consequently a maximum in pressure distribution ("Petrusevich pressure peak"), see Figure 2-29. Simultaneous solutions of the Reynolds equation and elasticity equations for line contact result in a pressure distribution and contact shape of the form illustrated by Figure 2-29. Increasing speed or viscosity causes the pressure spike to increase in height and to move from the outlet region toward the inlet. Dowson and Higginson (1959) developed an empirical minimum film thickness formula for lubricated line contacts as:

$$h_{min} = 2.65G^{0.54}U^{0.7}W_L^{-0.13}R \tag{2-91}$$

which predicts minimum films of about 1 μm; thus the gears operate in a lubrication regime in which significant interactions between surface asperities take place.

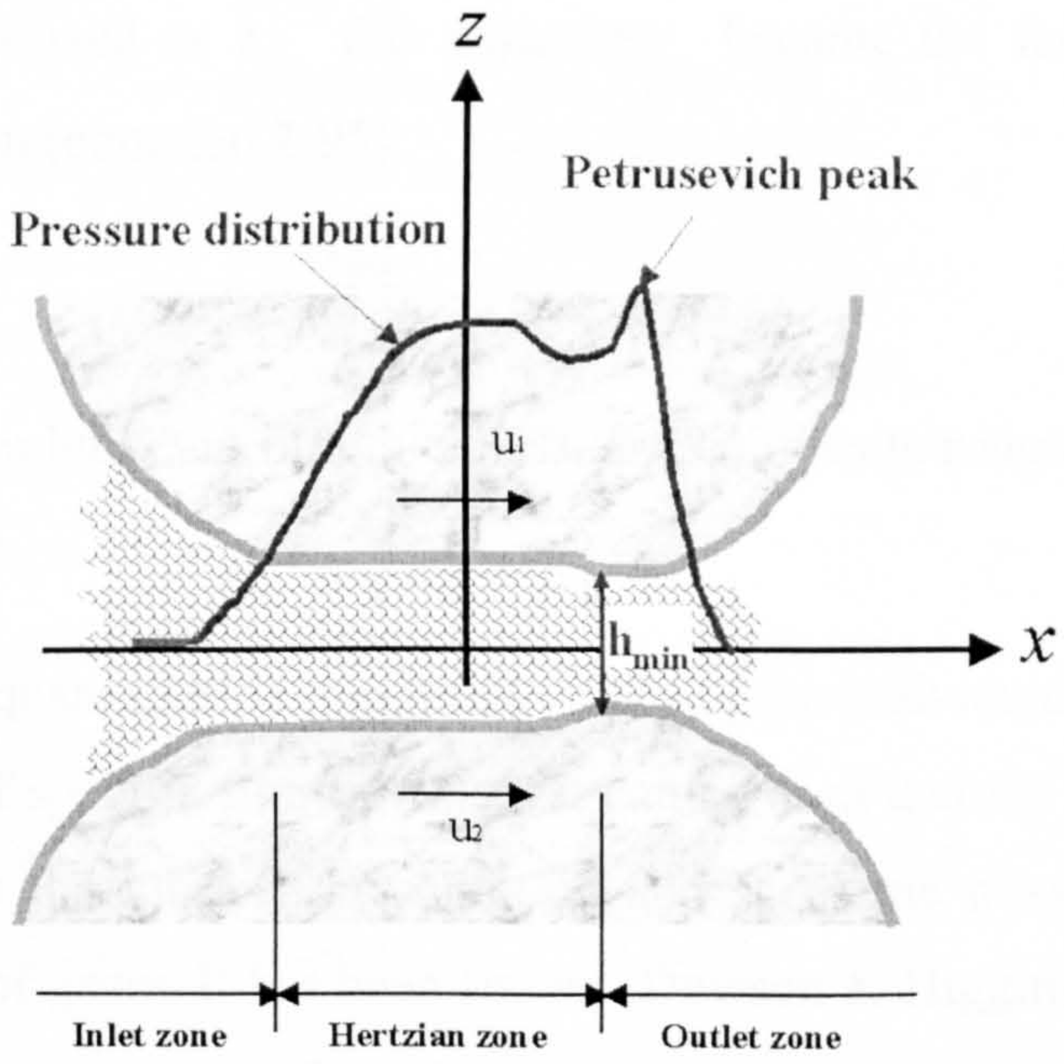


Figure 2-29. Pressure distribution and contact shape in EHL contact.

Gears exhibit an elliptical point contact rather than a line contact. A solution of the problem has been presented by Hamrock & Dowson (1976a, b & 1977). The resulting central and minimum film thickness are given by the following expressions:

$$h_{cen} = 2.69U^{0.67}G^{0.53}W^{-0.067}\left(1-0.61e^{-0.73k}\right)R_e \qquad (2\text{-}92)$$

$$h_{min} = 3.63U^{0.68}G^{0.49}W^{-0.073}\left(1-e^{-0.68k}\right)R_e \qquad (2\text{-}93)$$

where U and G have the same expressions as for line contact and W is given by equation (2-94).

$$W = \frac{P}{E' R_e^2} \qquad (2\text{-}94)$$

$k = (b/a)$ is the ellipticity ratio, a and b are the major and the minor semi-axis, respectively.

2.9.7 Specific film thickness (lamda ratio)

The film thickness itself does not give enough information in predicting surface fatigue life. In order to estimate the pitting life a relationship between lubricant film thickness and surface roughness has been introduced by Dawson (1962). Named *the specific film thickness* and symbolized as λ, this parameter became the factor used in surface fatigue life prediction (equation 2-95).

$$\lambda = \frac{h_{min}}{R_q} \qquad (2\text{-}95)$$

where: h_{min} minimum lubricant film thickness, R_q* composite roughness given by:

$$R_q{}^* = \sqrt{R_{q1}^2 + R_{q2}^2} \qquad (2\text{-}96)$$

R_{q1}, R_{q2} root mean square (RMS) roughness values of the individual surfaces.

From the formula of specific film thickness it can be seen that the reduction of surface roughness improves the lubrication condition and gives the possibility of increasing surface fatigue life of gears. It has been shown (Dowson & Higginson, 1977, Jacobson, 2000 and others) that lubricant film thickness needs to be about one to three times higher than the composite surface roughness in order to get long life of gears. Even so, according to Jacobson (2000) lambda values as high as 20 can give asperity contacts and lambda as low as 0.3 can stop metallic contact. It all depends on the local rheology

of the lubricant, surface roughness (i.e. slope and wavelength), contact pressure, rolling and sliding speed. Surface roughness will change until the surfaces almost totally conform to each other. Therefore it can be concluded that lambda ratio provides only a broad guide to the lubrication conditions (Jacobson, 2000).

The problem of lubricant film thickness has been satisfactory solved. The Dowson-Higginson equation (2-91) for line contact and Hamrock-Dowson equation (2-92 & 93) for elliptical point contact have been widely used in the design and analysis of machine elements although formulae that take into account other phenomena encountered in EHL contacts (e.g. side stravation) have been proposed: Sharif et al (2000). According to Bartz (2000) only about one third of the overall friction losses occur in the mixed-film or boundary regime, whereas two thirds are fluid film friction losses. Since rheological behaviour is highly relevant to automotive fuel efficiency, current interest is placed on the rheological properties of lubricants in EHL contacts.

2.9.8 Rheology of lubricants in EHL contacts

The rheological behaviour of lubricants in the very severe conditions present in elastohydrodynamic contacts is extremely complex. The lubricant in an EHD contact is subject to rapid and large pressure variations, large temperature changes and rapid transit time (Olver & Spikes, 1998). Under these severe conditions the assumption that lubricant has Newtonian behaviour fails. For these reasons lubricant rheology is the subject of great current interest. Comprehensive reviews of the theories that are used today to describe the rheological behaviour of lubricants in EHL contacts have been published by Moore (1997) and Dowson & Ehret (1999). A non-linear viscoelastic model that is consistent with all of the behaviour forms observed experimentally has been proposed by Evans and Johnson cited by Moore (1997). The frictional behaviour is divided into four regimes as shown in Figure 2-30.

- *Newtonian viscous behaviour* (I). In regime I the lubricant behaves as a Newtonian liquid and shear stress rises in proportion to shear rate. Viscosity follows the Barus equation.

- *Non-Newtonian viscous behaviour (Eyering)* (II). In Eyering regime the lubricant behaves intrinsically non-linear.
- *Viscoelastic behaviour* (III). In regime III it is assumed that lubricant responds to the applied shear rate with viscous and elastic displacements.
- *Elastic-plastic behaviour* (IV). In regime IV the friction coefficient tends towards a limiting value under the high pressure, which leads to the assumption that the lubricant solidifies under pressure and yield as a plastic solid. This means that a viscous fluid would approach a limiting value of friction coefficient at high pressure.

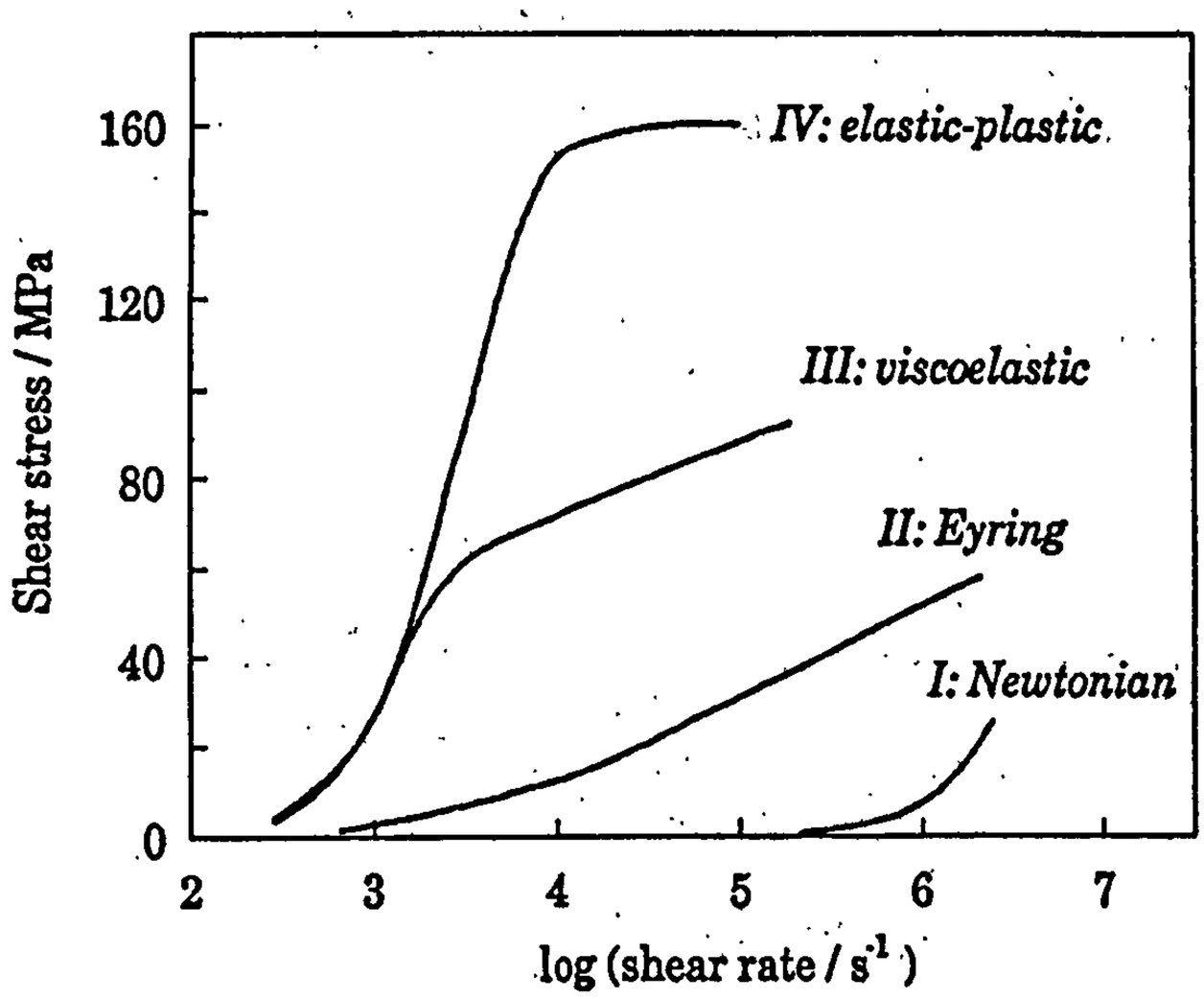


Figure 2-30. Evans and Johnson's model showing the four regimes of friction in elastohydrodynamic contact (from Moore, 1997).

2.9.9 Friction and temperature in EHL contacts

In rolling/sliding elastohydrodynamic contacts there is a strong dependence between temperature and friction. At the beginning, the friction determines the temperature of the lubricant film and also the heat conducted into the surrounding bodies. At the same time the lubricant film temperature determines the rheological properties and thus the friction. The mean coefficient of friction in a gear tooth contact can be expressed by equation 2-97 proposed by Smeeth and Spikes cited by Olver & Spikes (1998).

$$\mu = \mu_f + \frac{\mu_b - \mu_f}{(1+\lambda)^m} \tag{2-97}$$

where: μ_f is the fluid friction coefficient, μ_b is the coefficient of boundary friction ($\mu_b \cong 0.15$), m is a constant exponent, ($m \approx 2$), and λ is the specific film thickness.

A method to calculate the fluid friction coefficient has been proposed by Olver & Spikes (1998).

$$\mu_f = \frac{\tau^* \tau_E}{P} \tag{2-98}$$

where τ_E is the Eyring stress, $\tau^* = (\tau/\tau_E)$ is the non-dimensional shear stress given by the ratio of the shear stress, τ, to the Eyring stress, τ_E, and P is the mean contact pressure.

The shear stress, τ^*, has different expressions depending on the lubricant behaviour. If the behaviour is viscoelastic the formula for shear stress, proposed by Olver & Spikes (1998) is:

$$\tau^* \approx \ln(2S) - \ln\left(1 + 2\left|S - \frac{1}{2}\right| e^{-S/(2D_0)}\right) \tag{2-99}$$

where S is the non-dimensional shear rate

$$S = \frac{\gamma \eta_0}{\tau_E} \tag{2-100}$$

The shear rate is taken to be the nominal velocity gradient:

$$\gamma = \frac{U_s}{h_{cen}} \tag{2-101}$$

D_0 is the Deborah number - the ratio of the relaxation time of the fluid to the time of transit of the fluid through the contact.

$$D_0 = \frac{\eta_0 U}{2aG'} \tag{2-102}$$

γ is the shear rate, η_0 is the dynamic viscosity at low shear rate, U is the entrainment speed, a is the Hertzian contact half-width on entrainment direction, and G' is the effective elastic shear modulus given by:

$$\frac{1}{G'} = \frac{1}{G} + \frac{2a}{E' h_{cen}} \tag{2-103}$$

Owing to friction the surfaces of the two mating gears are heated. Surface tooth contact temperature is given by the sum of bulk temperature and the "flash temperature". The

temperature rises due to dissipation of frictional heat at the peaks of the contacting asperities. The study of asperity flash temperature is important in partial-EHL lubrication where the temperature rise reduces the lubricant adsorbate that protects the surfaces from scuffing. It is difficult to accurately measure the flash temperature therefore theoretical analyses have been used for its prediction. Following the pioneering work of Blok (1937) many other researchers have proposed methods for calculating the flash temperature: Jaegar, 1942; Archard, 1959; Guha & Roy Chowdhuri, 1996; Abdel-Aal, 1997; Pandey & Ghosh, 1998, Gao et al., 2000).

Olver & Spikes (1998) proposed a formula for the mean temperature of the lubricant within the contact as:

$$\overline{T} = T_B + \Delta T_f + \Delta T_{lub} \tag{2-104}$$

where: $\overline{T}$ average temperature of the lubricant film within the contact, T_B average temperature of the lubricant film in the inlet region, ΔT_f average flash temperature, ΔT_{lub} internal heating of the film itself. They are given by the following equations.

$$\Delta T_f = 1.06 B_1 \alpha_h q \tag{2-105}$$

$$\Delta T_{lub} = \frac{U_s \bar{\tau} h_c}{8k} \tag{2-106}$$

where, B_1 is the transient thermal resistance, α_h is the proportion of total heat entering body 1, q is the total heat generated in the contact, U_s is the sliding speed and k is the thermal conductivity of the lubricant.

In a paper by Castro and Seabra (1998) describing scuffing in an FZG gear test the values reported for the surface tooth contact temperature at the scuffing stage were 142, 225 and 415°C for three different types of base oils, ISO VG68, 220 and 680 respectively. The overall conclusion is that the tooth surface temperature significantly increases as the viscosity of the base oil increases.

2.9.10 Lubrication of rough surfaces

Gear tooth contact often operates in the regime of micro-EHL and boundary lubrication due to thermal effects, non-Newtonian lubricant effects and surface roughness. The surface finish of most gears is about R_a = 0.4 μm and lubricant film thickness is about

1μm or, far less under high-temperature conditions. Under these conditions a certain amount of metallic contact will take place. In gears the surface roughness features (i.e. grinding marks) are perpendicular to the direction of rolling and sliding, therefore perpendicular to the direction of lubricant entrainment. In these conditions there is an opportunity for a loss of oil from the contact which leads to thinning of the film and ultimately to the film collapse. According to Evans & Snidle (1996) such conditions occur at the edges of real contacts in the valleys where pressure and viscosity are low. Due to local pressure variations within the contact lubricant leaks into the surface valleys allowing the asperity tops to reach through the lubricant film and colide. Thus, the interaction between asperities and the lubricant rheology determines the lubrication conditions (Jacobson, 2000). A review of the work done on the influence of surface roughness features upon the EHL conditions has been published by Greenwood (1999). There is a good understanding of the behaviour of Newtonian fluids, however non-Newtonian behaviour is still disputed.

The roughness inside an EHL contact is different to that outside the contact owing to elastic deformation. Elastic deformation of asperities disturbs the Hertzian pressure distribution within the contact.

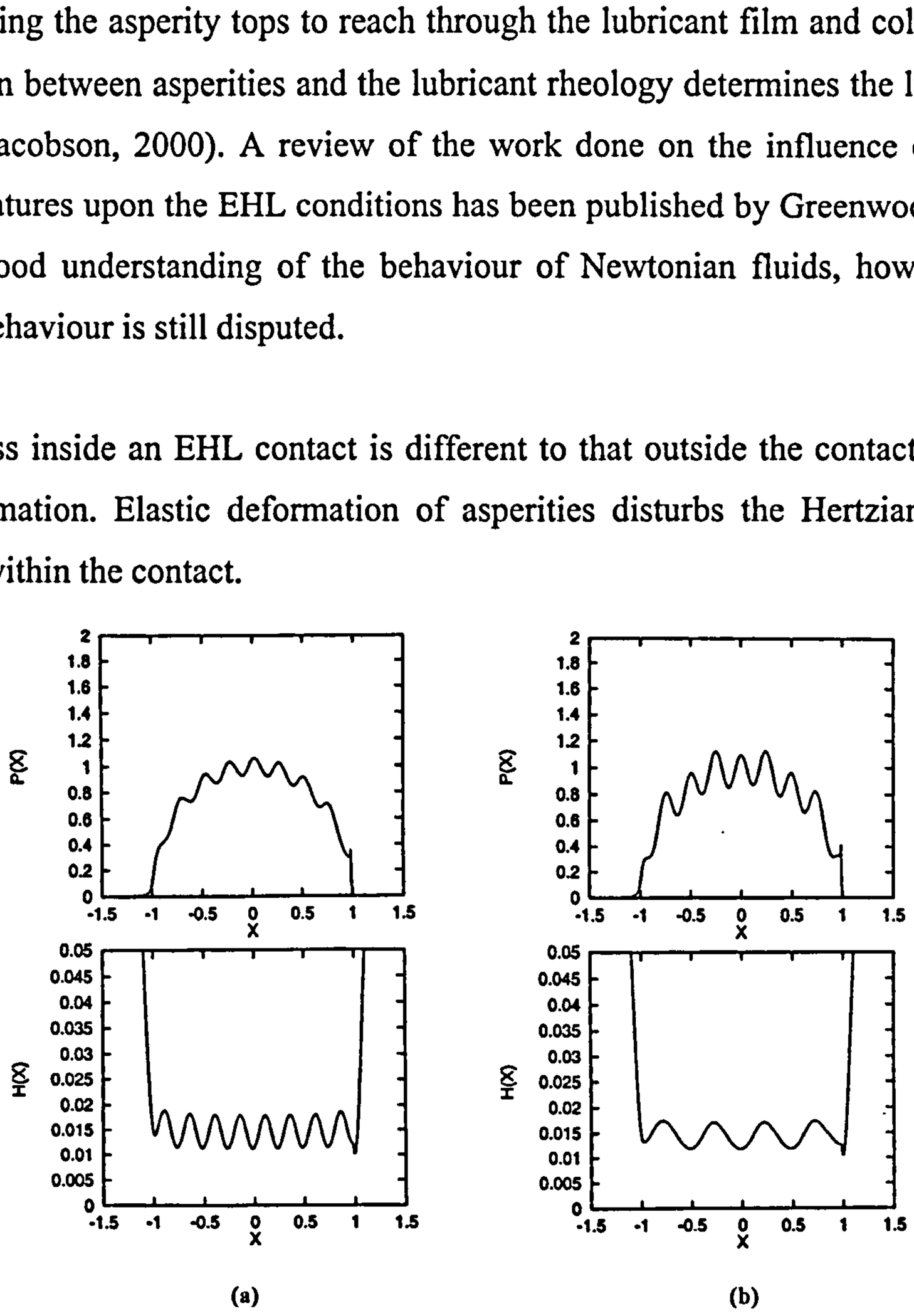


Figure 2-31. Pressure and film thickness distribution in a line contact: (a) under pure rolling, (b) under rolling/sliding (from Lubrecht & Venner, 1999).

Lubrecht & Venner (1999) pointed out that unlike in pure rolling, where the pressure fluctuations and the film thickness variations are in phase, in rolling/sliding contacts pressure and film thickness perturbations have different wavelengths, see Figure 2-31, which makes the study even more complicated.

A challenging hypothesis, which is contrary to the current EHL theory, has been made by Guangteng et al (2000). They investigated the influence of different roughness features (transverse and longitudinal ridges, circular bumps) in condition of $0.1<\lambda<2$. All of these features are heavily deformed and progressively recovered as the rolling speed increases. They found that the local pressure field at the asperities becomes more severe at high speeds and film thicknesses than at low ones. This is contrary to the current EHD theory in which the magnitude of the pressure field is determined by the applied load so that once the pressurised film develops, asperity pressure falls. The reason for this is not yet clear and further work is needed to verify the phenomenon.

Although the effects of surface irregularities on rolling/sliding EHL contacts have been amply investigated in the last few years it is still difficult to provide a conclusive theory. It must be taken into account that the lubricant in EHL contact behaves as an elastic-plastic and visco-elastic solid rather than a viscous fluid (Kaneta & Nishikawa, 1999). Furthermore, EHL films are affected not only by surface irregularities but also by air bubbles. Over the past thirty years an astonishing progress in numerical solutions of the rough surface elastohydrodynamic problem has been achieved. Lubrecht & Venner (1999) emphasised that the features present in a roughness profile have a very small scale of wavelength, comparable with the distance between two measuring points (~1μm). In order to represent such features with sufficient accuracy a mesh size of 10^{-8} m would be required. The computation using today's generation of fast algorithms seem to be beyond the capacity of the current series of workstations. The solution of a problem with 10^{10} points and 10^5 time steps may remain impossible for a long time to come.

2.9.11 Gear lubricants

The basic functions of a lubricant are friction reduction, heat removal and suspension of contaminants. A gear lubricant usually consists of a base fluid, generally of petroleum origin, most often combined with additive chemicals that enhance the properties of the base fluid. Base fluids are essentially obtained from two main sources: the refining of petroleum crude oil and the synthesis of relatively pure compounds with properties that are suitable for lubricants. Synthetic lubricants provide some advantages compared to petroleum-based fluids, such as a better thermal stability, superior volatility characteristics, more desirable frictional properties, and an improved viscosity/temperature characteristics.

2.9.11.1 Lubricants characteristics

The selection of a lubricant for a given application is made in conjunction with several characteristics. In what follows a brief description of the most important characteristics is given.

Flash point is a measure of lubricant volatility and represents the temperature at which an oil vaporises sufficiently to sustain momentary ignition when exposed to a flame under atmospheric conditions. Volatility is essential in the choice of lubricants and cannot be improved by the use of additives. A lubricant must exhibit low volatility under operating conditions.

Pour point is the lowest temperature at which an oil will flow. This characteristic can be improved by the use of additives called pour point depressants.

Cloud point is the temperature at which wax precipitation starts.

Viscosity index (VI) is an arbitrary number used to characterise the variation of kinematic viscosity of an oil with temperature. Additives, called viscosity modifiers can improve viscosity, at high temperatures.

Oxidation stability is the resistance to a reaction with oxygen.

2.9.11.2 Mineral oils

Mineral oils are processed from crude oil, which has had coarse impurities, petroleum gas, and water removed. The main constituents of crude oil are hydrocarbons. The basic types of hydrocarbon mixtures that form a mineral oil can be classified as follows (Bartz, 1993).

Chain-like or aliphatic, saturated hydrocarbons. Their collective names are alkanes or paraffins (C_nH_{2n+2}).

Chain-like, unsaturated hydrocarbons also known as alkenes or olefins (C_nH_{2n} and C_nH_{2n-2}).

Cyclic, saturated hydrocarbons, with the collective names cycloalkanes, cycloparaffins or naphthenes (C_nH_{2n}).

Cyclic, unsaturated hydrocarbons with the collective name cycloolefins (C_nH_{2n-2}) and aromatics (C_nH_n).

2.9.11.3 Synthetic oils

Synthetic oils are developed for use in special applications (e. g. extremely high or low temperatures and loads). Synthetic oils of commercial interest are (Bartz, 1993): synthetic hydrocarbons (*polyalphaolefins* and *alkyl benzenes*), *polyglycols*, *esters*, and *silicone lubricants*.

The characteristics of synthetic lubricants derive from the physical and chemical characteristics of the base oil and added substances. Polialphaolefins have a high viscosity index, excellent low-temperature performance and low volatility. Polyglycols have also a very low pour point. Alkyl benzenes can be manufactured in almost all viscosity grades. Esters have good high-temperature and low-temperature flow characteristics. The best scoring resistance as reported by Bartz (1993) was obtained with polyalphaolefins and esters and good friction performance with polyglycols.

2.9.11.4 Lubricant additives

The viscosity of any fluid changes with temperature, increasing as temperature decreases. Changes in thermal and oxidative stability of a lubricant will decrease its

efficiency in heat removal. Additives are generally employed to solve such problems. The additives enhance some properties of the lubricant but some others are negatively influenced (e. g. EP additives in gear oil increases the extreme-pressure capabilities but decreases the oxidation life). Generally, a lubricant contains more than one additive. The most common additives are as follows:

Antioxidants have the role of decreasing the rate of oil oxidation. Oxidation increases the viscosity of lubricant and results in formation of peroxides and organic acids. Antioxidant additives reduce the tendency of hydrocarbons to combine with oxygen.

Corrosion inhibitors are substances that form protective films on metallic surfaces and thus stop direct contact between gear tooth surface and aggressive components found in lubricant.

Pour-Point Depressants are high-molecular-weight polymers, which inhibit the formation of wax crystals within the oil. The tendency of wax formation is higher in oils containing paraffinic hydrocarbons.

Viscosity Index Improvers increase the resistance of oil to thinning with increasing temperature. Their effect is several times greater at high temperatures than at low temperatures, and thus causes a reduction in the temperature-viscosity dependence.

Antifoam agents reduce the surface tension of the oil, allowing the air bubbles to combine into larger bubbles and then dissipate into the air.

Antiwear (AW) Additives are used where metal-to-metal contact is present (e. g. boundary lubrication). AW additives react chemically or chemically and physically with the flank surfaces and form protective layers. These layers are continually removed and reformed so metal to metal contact does not take place.

Extreme Pressure (EP) Additives are used for preventing high or shock loading of the components. EP additives are activated by heat rather than pressure. They react with steel surface and form a protecting film or compound.

Of a particular importance in gear lubrication are antiwear (AW) and extreme-pressure (EP) additives, which are activated only under specific temperature and pressure conditions and are inactive under other conditions.

EP and AW additives can be classified as *chemically active additives* and *physically active additives*. Chemically active additives include *sulphur compounds* (sulphurated fat oils, sulphurated terpenes, olefins and dibenzyl disulphides), *phosphorus compounds*

(tricresyl phosphate), *chlorine compounds* (chloroparaffin), *lead compounds* (naphthenates and lead oleates) or *mixed compounds* (dialkyl dithiophosphates based on zinc). Physically active additives act by adsorbtion to the metal surface. Included in this group are soaps, etsers, fatty oils, and organic acids.

AW additives function in moderate environments of temperature and pressure while EP additives are effective in the more extreme environments.

As far as the fuel economy of passenger cars is concerned, according to Bartz (2000) only 12 % of the input energy is converted into usable energy (energy at the wheels to overcome the tyre and air resistance). Bartz emphasised that only one-third of the friction losses occurs in the mixed-film or boundary regime, whereas two-thirds are fluid-film friction losses. According to Bartz cited by Bala et al. (2000) up to 6.3 % improvement can be realised by the use of lower viscosity multigrade gear lubricants containing friction modifiers. Gear lubricants with high VIs promote mixed lubrication regime between contacting gear teeth. The ideal gear lubricant thus has to have its viscosity-temperature behaviour optimised along with its additive system (Bala et al, 2000). In order to improve fuel economy a gear lubricant operating at low temperatures must have low viscosity to maintain high torque efficiency. At higher operating temperatures, the recommended lubricants are those combining a high VI and boundary sulphur-phosphorus additive.

2.10 Summary

This chapter provides the theoretical background concerning gear manufacturing, heat treatments and the microstructure achieved in case carburised steel. It also provides information about stresses in gears, surface roughness and gear lubrication. This information is sufficient to understand how gear can fail during operation.

The failure modes are presented in the next chapter, emphasis being placed on micropitting failure, which is the main subject of this work.

~ Chapter 3 ~

MICROPITTING AND RELATED PHENOMENA

3 MICROPITTING AND RELATED PHENOMENA

3.1 General

Micropitting is a form of surface contact fatigue failure which appears in many applications of lubricated gears (i.e., heavily loaded gearboxes with surface hardened-teeth including wind turbines, plastic extruders, and gearboxes found in the paper, steel and cement industries). Micropitting can occur after a relatively short period of operation and, after a certain number of cycles, gears damaged by micropitting need to be replaced due to the increased noise and vibrations caused by the deviations of the tooth profile. Continuing operation of affected gears can lead to a catastrophic type of failure (i.e., tooth breakage). Replacing gears requires down times and extra costs or, an even worse scenario, gear failure can have safety implications (e.g., aircraft industry). These considerations explain the increasing current interest in micropitting. Previous research has shown that micropitting is associated with microstructural transformations in steel. However, to the author's knowledge, these transformations have not been reported in gears but only in rolling bearings. In the present work extensive metallurgical investigations have been carried out and they have revealed the same transformations occur in gears.

This chapter deals with gear failure modes, emphasis being placed on failure due to surface contact fatigue. A distinct section is dedicated to the material response to surface cyclic loading which includes the microstructural transformations. The features that are specific to micropitting, the mechanisms involved and the factors of influence, as they have been proposed in literature are extensively presented. Methods for micropitting prevention are also reviewed.

3.2 Gear failures classification

A failed gear can be defined as a gear that can no longer efficiently perform the job for which it was designed (Shipley, 1974). In general, each failure mode can be deduced by investigating the condition of the teeth after failure. Gear failures can be classified in the following general classes (ASM, 1975 and BS 7848: 1996): *wear*, *scuffing*, *plastic flow*, *breakage*, and *surface fatigue*.

Wear is defined as the loss of material from contacting surfaces of the teeth. It is further classified as *normal wear*, *moderate wear*, *abrasive wear*, *scoring*, *interference wear*, and *corrosive wear*.

Scuffing consists in localised welding of the tooth flanks as a result of the disruption in the lubricant film. Severe scuffing causes an increase in noise and vibration and rapidly leads to failure.

Plastic flow is surface deformation resulting from yielding of surface metal under heavy loads. It is further classified as *rolling and peening*, *rippling*, and *ridging*.

Breakage is defined as fracture of an entire tooth or of a substantial portion of a tooth. It is further classified as *fatigue breakage*, *breakage from heavy wear*, *overload breakage*, *quenching cracks* and *grinding cracks*.

Surface contact fatigue is the failure of a gear as a result of repeated surface or subsurfaces stresses that exceed the endurance limit of the steel. It is further classified as *pitting*, *flaking*, *spalling*, *case crushing*, and *micropitting*.

As it has been already shown, the shape of gear tooth contact becomes a small ellipse when the load is applied. The area of contact is in compression and the displacements of the metal result in subsurface shear and tensile stresses. Frictional forces added to the stresses impose additional surface compressive and tensile stresses. Therefore, the material at and near the surface is subject to compression, tension and shear every time it passes through the contact. The repeated stresses may lead to the initiation and propagation of a fatigue crack. Surface fatigue phenomena play a significant role in gear failures. Figure 3-1 shows a comparison of gear failure modes. Two or more failure modes may occur simultaneously, or one may be the result of the other (ASM, 1975). For example, a microcrack associated with micropitting may grow into a macrocrack

and propagate until it causes the fracture of the entire tooth. The above classification is meant to assist in distinguishing between cause and effect and in determining the appropriate corrective action (ASM, 1975).

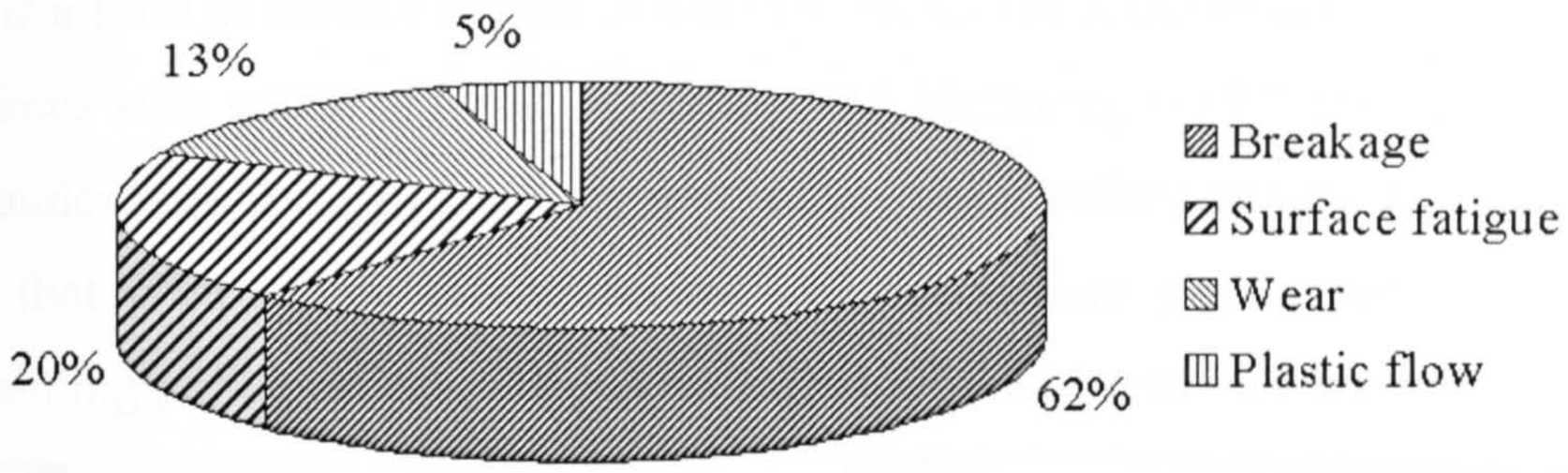


Figure 3-1. Comparison of gear failure modes. Damages due to surface contact fatigue take place in 20 % of the cases (after ASM, 1975).

3.3 Surface contact fatigue

Surface fatigue failure has been defined by BS 7848: 1996 as "material damage due to surface and subsurface stresses produced by the repeated application of forces. It is characterised by the removal of metal and the formation of cavities". The term "fatigue", which means tiredness in French, has been mentioned for the first time in the English language in 1867, in an eulogistic article published anonymously in *Engineering*, about Wöhler's exhibit at the Paris exhibition of that year (Fine, 1980). Wöhler, the most prominent name in the history of fatigue, established the first fatigue laboratory and carried out the first systematic fatigue tests. In the old days fatigue was described as a disease of metals which occur by "crystallisation" or molecular changes in microstructure. Rankine, cited by Fine (1980), was the first to recognise in print, in 1843, that fatigue occurs by *crack initiation and propagation*. The first metallographic study of fatigue has been published by Ewing & Humfrey (1903), cited by Fine (1980). They have shown that fatigue cracks form along *slip bands* (see section 3.3.2).

3.3.1 Classification of surface contact fatigue phenomena

Surface fatigue phenomena are classified as follows (BS 7848: 1996).

Pitting is a form of surface fatigue in which particles break out of affected areas leaving the surfaces with scattered holes. Fernandes and McDuling (1997) distinguish between pitting under pure rolling conditions and pitting under rolling/sliding conditions. They believe that pure rolling pitting occurs with no surface plastic deformation while rolling/sliding pitting occurs with plastic deformation of the surface material.

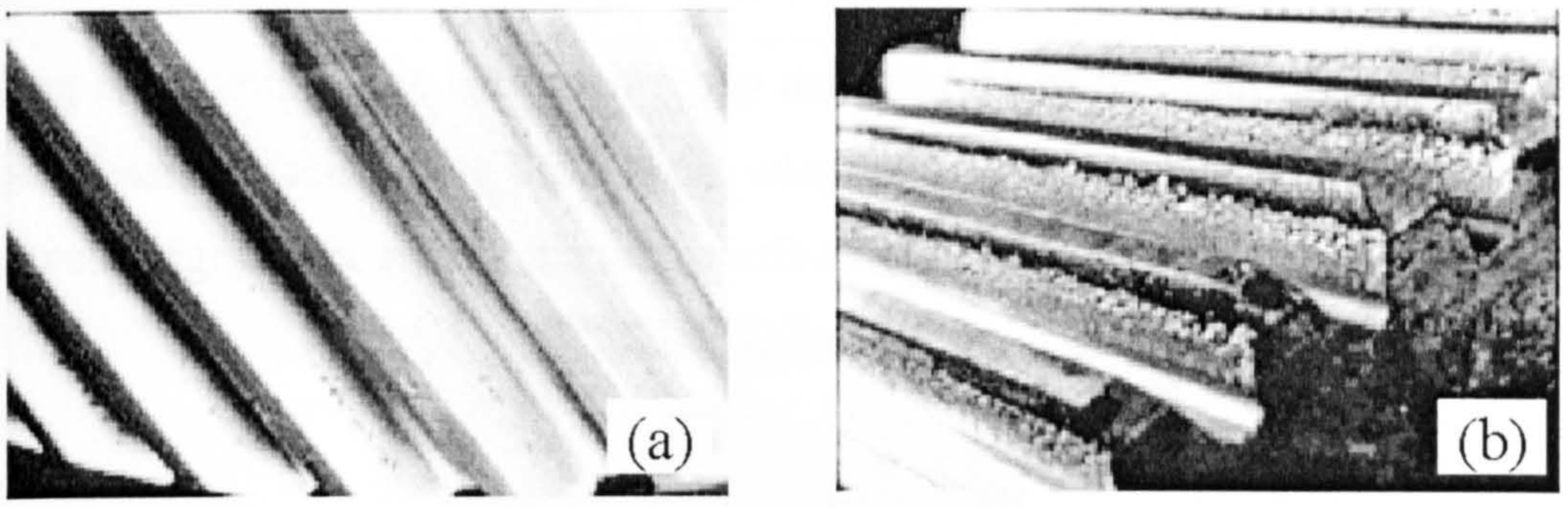


Figure 3-2. (a) Initial pitting and (b) progressive pitting (from BS 7848: 1996).

Initial pitting presents areas of small pits from 300 μm to 800 μm in diameter, which occur in localised parts of the gear teeth that are over-stressed. It tends to redistributes the load by progressively removing high contact spots, and often stops once the load has been redistributed. Continued operation may polish the pitted surface and improve its appearance (see Figure 3-2a). This pitting originates at the surface (Graham, 1961).

Progressive pitting is the pitting which continues and which tends to increase throughout the gear life (see Figure 3-2b). It appears as much larger pits than initial pitting, often in the dedendum section of the gear tooth. These larger pits are usually caused by more severe overload conditions that cannot be relieved by initial pitting. Progressive pitting may originate from either subsurface or surface initiated cracks (Graham, 1961).

Flaking is a form of tooth surface damage involving the braking out of thin flakes of material of comparatively large area leaving shallow cavities of roughly constant depth shaped like inverted triangles.

Figure 3-3. Flaking (from www.tribology.co.uk)

Spalling is sometimes the term incorrectly used instead of the term of flaking. Spalling is a progressive macropitting that occurs when pits coalesce and form irregular craters that cover a significant area of the tooth surface.

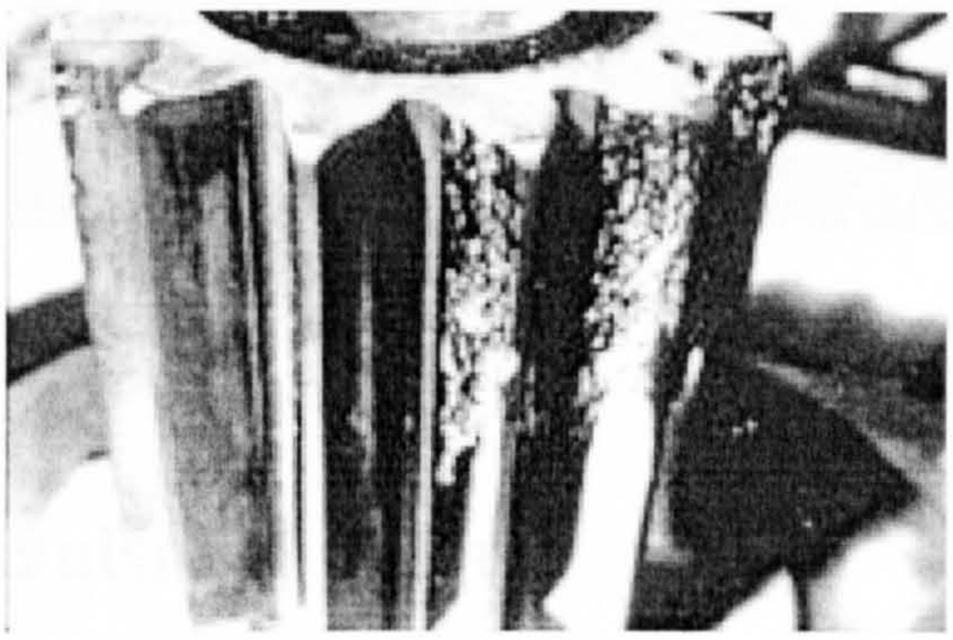


Figure 3-4. Spalling (after BS 7848: 1996).

Case crushing resembles progressive pitting, except that the pits may be larger, quite shallow, and irregularly shaped. The edges of the pits break away rapidly, forming large, irregular voids that may join together. Material is detached from the surface due to propagation of cracks, usually along the case/core transition zone.

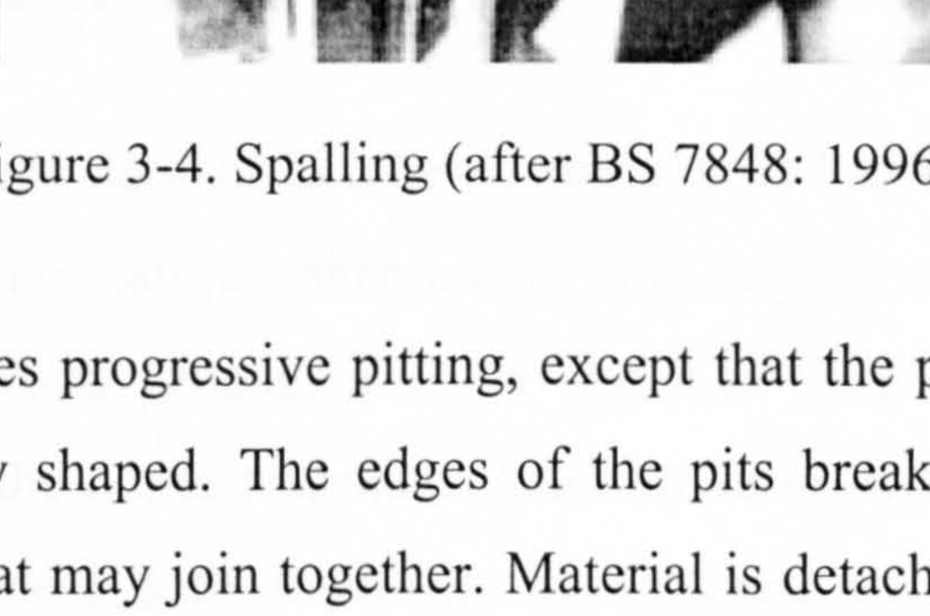


Figure 3-5. Case crushing (after De Lange, 2000).

Micropitting, also known as *frosting* or *grey staining*, is a type of contact fatigue damage that appears under thin film lubrication conditions. The surface acquires an etch-like finish with a pattern that sometimes follows the slightly higher ridges left by grinding marks or other surface irregularities. The surface appears as a field of very fine micropits of 1-30 μm deep. It usually starts to appear on the dedendum section of the driving gear, although it may begin on the addendum section as well. Causes include high surface loads and heat generation, which thin the lubrication film and leads to boundary lubrication.

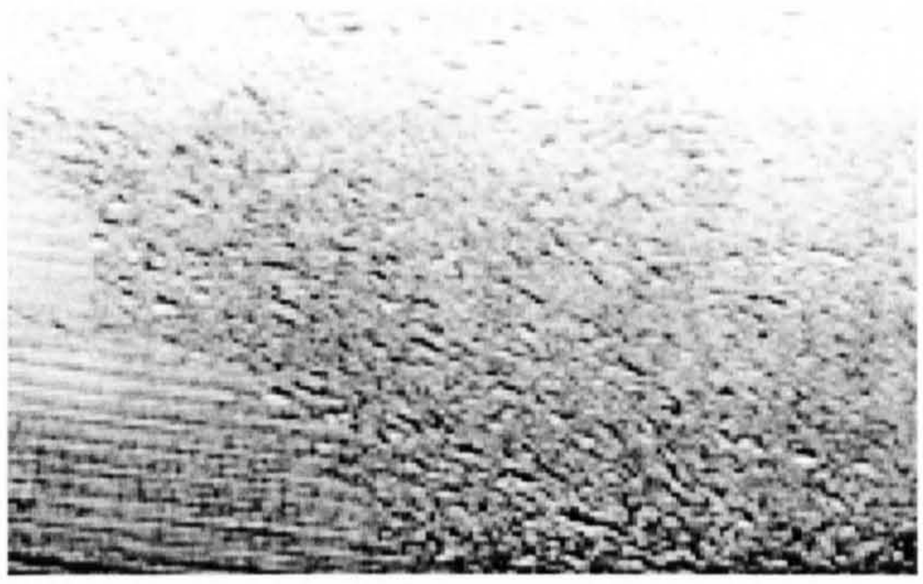


Figure 3-6. Micropitting (after BS 7848: 1996).

3.3.2 The mechanism of fatigue

The mechanism of surface contact fatigue phenomena is based on crack initiation and propagation as in all other fatigue phenomena. *Sub-surface cracks* tend to initiate at non-metallic inclusions and propagate at an angle of 20-30° to the surface in the direction of motion of the contact surface. After some growth the crack may branch. Crack branching may occur both down into the material and up towards the surface. When the crack branch reaches the surface it causes the removal of wear particles.

Surface cracks initiate at the component surface at sites of damage or extremes of roughness. The asperities act as stress raisers and cause the formation of cracks. The crack propagates at an angle about 20° to the surface then turns parallel to the surface before at some point branching occurs and a pit is formed. Surface damage can also be caused by the entrainment of lubricant debris particles. Typically the initiation phase takes between 10 and 40 % of the total time to failure (Roylance et al., 2000). Several stages may be distinguished in the process of any fatigue failure: *cyclic plastic*

deformation (i.e., long range dislocation motion*), initiation of microcracks*, *propagation of microcracks*, *propagation of macrocracks*, and eventually the *failure* (Fine, 1980).

3.3.2.1 Cyclic plastic deformation

About 90 % of the failures of gears occur at stresses below their proportional limit. In uniaxial loading, ductile materials begin to deform plastically at stresses equal to a threshold known as the *yield stress*. There are two generally accepted criteria for predicting the onset of yielding: *Tresca criterion*, also known as *maximum shear stress criterion* and, *von Mises criterion*, also known as *distortion energy criterion*.

3.3.2.1.1 Tresca yielding criterion

The Tresca yielding theory states that yielding occur when the maximum shear stress reaches a critical value equal to the shearing yield stress in a uniaxial tension test.
The maximum shear stress is given by

$$\tau_{max} = \frac{\sigma_1 - \sigma_3}{2} \tag{3-1}$$

where, σ_1 is the maximum and σ_3 is the minimum principal stress.
For uniaxial tension $\sigma_1 = \sigma_0$, $\sigma_2 = \sigma_3 = 0$, where σ_0 is the yield strength in simple tension. Therefore, the shearing yield stress, τ_0, for simple tension is

$$\tau_0 = \frac{\sigma_0}{2} \tag{3-2}$$

Then the yielding criterion becomes

$$\tau_{max} = \tau_0 = \frac{\sigma_0}{2} \tag{3-3}$$

3.3.2.1.2 von Mises yielding criterion

According to the von Mises criterion, yielding will occur when the differences between the principal stresses exceed the yield stress in uniaxial tension, σ_0. The mathematical form of von Mises criterion is given by the equation below

$$\sigma_0 = \frac{1}{\sqrt{2}}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right]^{1/2} \quad (3\text{-}4)$$

The von Mises' criterion predicts that the yield stress in torsion will be less than in uniaxial tension according to equation 3-5.

$$\tau_0 = \frac{\sigma_0}{\sqrt{3}} = 0.577\sigma_0 \quad (3\text{-}5)$$

During alternating stress, plastic deformation occurs in localised areas and the dislocation density increases within grains. The dislocations may emerge at the surface and creates a slip step of one Burgers vector or they pile-up against obstacles (e.g. grain boundaries, inclusions, etc.). An accumulation of the slip steps at the surface forms a slip band and can be visualised in some cases in the form of *extrusions* and/or *intrusions* as indicated in Figure 3-7.

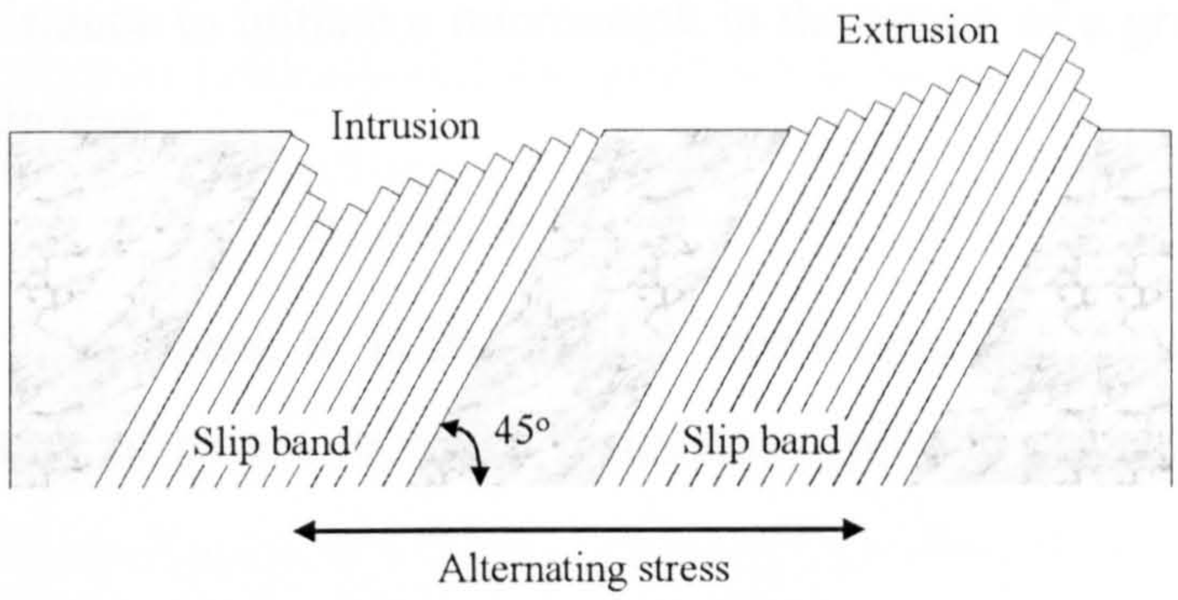


Figure 3-7. The formation of slip bands and intrusions and extrusions at a metallic surface occur at 45° to the direction of alternating stress.

The formation mechanism of extrusions/intrusions has been explained by the interaction of edge dislocations on two slip systems (Cottrell & Hull, 1957), the cross slip of screw dislocations (Mott, 1958), and by a dislocation avalanche along parallel slip planes containing dislocation pile-ups of opposite sign (Neumann, 1969). Thomson et al. (1956) observed that some slip bands reappeared in the same places after their removal by electropolishing and re-testing the specimens. They called these bands *persistent slip bands*. It is believed that extensive plastic deformation occurs in the persistent slip bands (Fine, 1980). Cyclic plastic deformation causes cyclic hardening or cyclic softening depending on the material. The quenched and tempered steels cyclically soften by as much as 40 % due to a change in dislocation structure caused by cycling (Fine, 1980). Tempering reduces the density of dislocations but many dislocations will

remain from the martensite transformation. Due to cycling straining the dislocation density is further reduced and coherent precipitates will form, which counteracts the cyclic softening to some extent.

3.3.2.2 Microcrack initiation

The preferential sites for microcrack initiation are the slip bands, the grain boundaries, the second phase particles (i.e., carbides), and the interface between inclusions and the second phase particles. As cycling continues, more and more dislocations emerge at the surface and the slip band widen and, ultimately, becomes a microcrack. Because the yield stress is lower in the centre of the grain this process is more intense in this region. Considering the Hall-Petch (Hall, 1951 & Petch, 1953) relation (Equation 3-6) it can be stated that the resistance to initiate a microcrack in the centre of a grain decreases with increasing the grain size.

$$\sigma_y = \sigma_i + KD^{-1/2} \tag{3-6}$$

where, σ_y is the yield stress, σ_i is the friction stress opposing dislocation motion, K represents a measure of extent to which dislocations pile-up at barriers and, D is the grain diameter.

The dislocation pile-ups may promote a dislocation avalanche if a critical value has been reached. This can lead to the formation of a microcrack in the matrix, along the grain boundary, along the interface between the matrix (i.e., martensite) and the second phase (i.e., carbides), or in a second phase particle (i.e., carbide). According to Fine (1980), if the ratio of Young's modulus of the particle to that of the matrix is larger than one (i.e, Al_2O_3 in Fe) then a large stress concentration will arise around the particle.

3.3.2.3 Microcrack propagation

There is no definition universally accepted of a microcrack. Researchers have defined a microcrack as a crack of an arbitrary length, usually crossing one or two grains and a macrocrack as a crack larger than several grain diameters. The persistent slip band deformation process, which forms an extrusion or intrusion and then a microcrack can lead to the propagation of that crack. Since the persistent slip bands are formed at 45° to

the stress axis (on the most highly stressed plane), the microcrack will propagate on this plane. *This is called the stage I crack propagation* (see Figure 3-8).

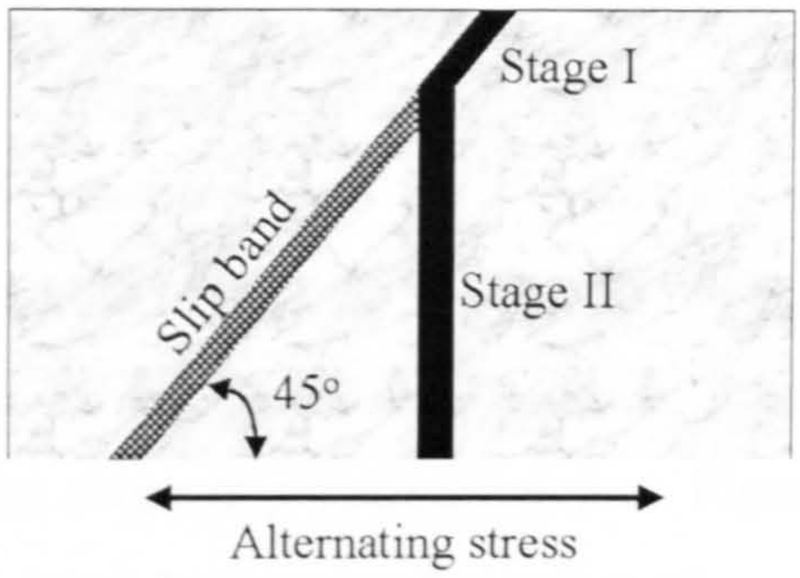


Figure 3-8. Diagram showing the crack propagation in two stages: stage I at 45°, and stage II normal to the stress axis.

Not all the microcracks are viable to propagate. It is well known that microcrack propagation is impeded by grain boundaries because plastic deformation is discontinuous across grain boundaries. A suitable criterion for crack propagation is that of Griffith (1920). Although Griffith's theory is applicable to perfectly brittle materials, it has had a great influence on explaining crack propagation in metals and alloys. The Griffith's criterion postulates that "A crack will propagate when the decrease in elastic strain energy is at least equal to the energy required to create the new crack surface."
The Griffith's criterion gives the stress, σ, required to propagate a crack as a function of its length, a (see Figure 3-9).

$$\sigma = \left(\frac{4E\gamma_s}{\pi a} \right)^{1/2} \tag{3-7}$$

where, E is the Young's modulus, and γ_s is the surface energy.

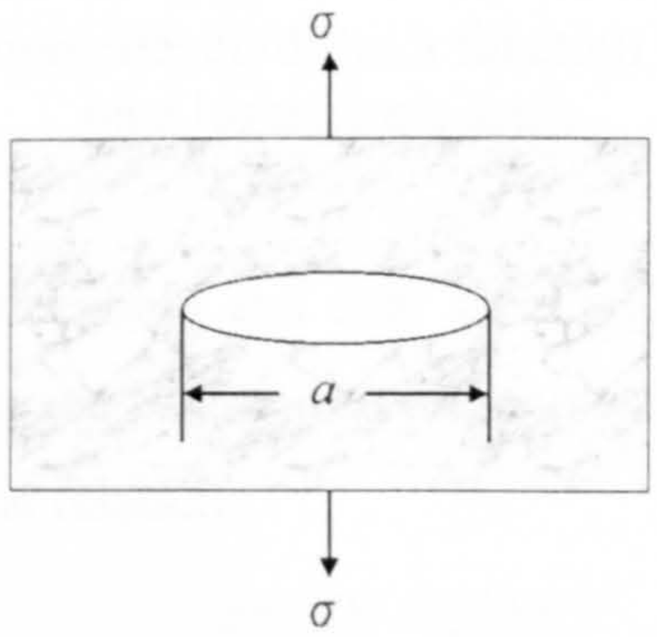


Figure 3-9. Elliptical crack model (Griffith).

3.3.2.4 Macrocrack propagation

A microcrack can propagate in stage I mode to a length of several grain diameters when it becomes a macrocrack. When the *stress intensity factor range*, ΔK, (see below) reaches a limit, the crack promotes slip on systems other than the primary and a dislocation structure forms at the crack tip leading to the break down of the persistent slip bands structure (Laird, 1979). The crack begins to propagate normal to the stress axis. This is known as the *stage II propagation* (see Figure 3-8). The rate of crack propagation in stage II is much higher (on the order of μm per cycle) compared with crack propagation rate in stage I (on the order of nm per cycle) (Dieter, 1988). The stress intensity factor range, ΔK, is defined as

$$\Delta K = (\sigma_{max} - \sigma_{min})\sqrt{a} f(g) \tag{3-8}$$

where σ_{max} and σ_{min} are the maximum and minimum nominal stress, respectively, a is the crack length and f(g) is a function of specimen geometry and loading conditions. The crack propagation rate (i.e., the crack length versus number of cycles), da/dN varies as shown in Figure 3-10.

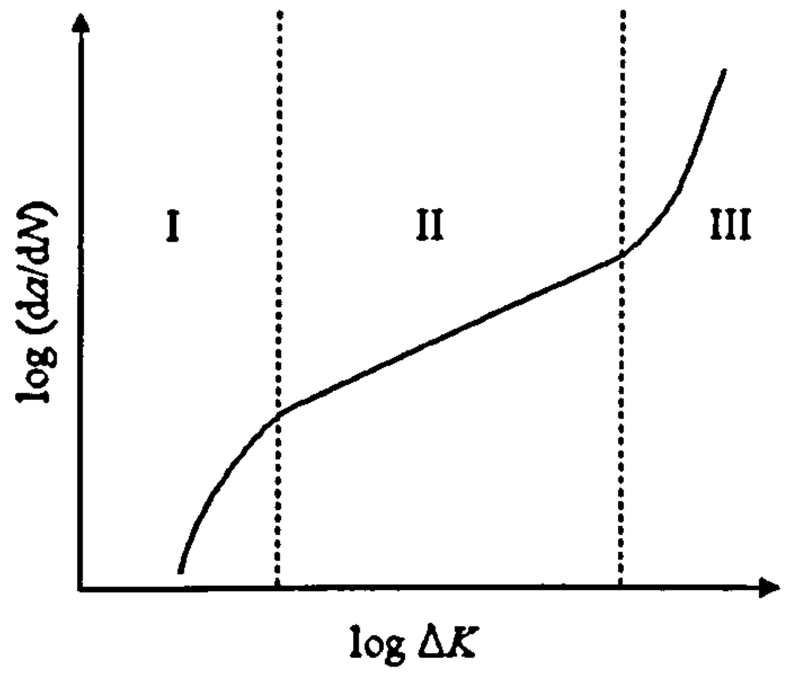


Figure 3-10. Crack propagation rate, da/dN as a function of stress intensity factor, ΔK, on a logarithmic scale.

Regime I characterises the threshold and near threshold region. At stresses below a threshold cracks do not propagate. Region II represents a linear relationship between log da/dN and log ΔK given by the relation

$$da/dN = C(\Delta K)^m \tag{3-9}$$

where C and m are constants. In region III the crack propagates very rapidly.

3.4 Material response to surface cyclic loading

During running, the surface of the gear tooth is cyclicly exposed to the load and, as a consequence, elastic and plastic conditions are periodically changed in the regions close to the surface. Friction between the contacting surfaces generates heat, which is also transmitted to the near-surface regions. As a result the microstructure of the steel below the running surface goes through to a "self healing" process (Bush et al., 1961) which leads to a series of phenomena occurring up to a depth usually corresponding to the depth of maximum shear stress. The phenomena concerned are as follows: changes of the residual stress pattern, microstructural transformations (transformation of retained austenite and martensite decay, the latter reported only in bearing to date), nucleation and propagation of microcracks.

3.4.1 Residual stress changes

As mentioned in Chapter 2, compressive residual stress in the near surface region can increase the fatigue life of gears by delaying the crack initiation and propagation. Under the influence of load and/or the heat generated by friction changes in the residual stress pattern can occur by dislocation movement, which converts the residual elastic strains into microplastic strains (Batista et al., 2000). Moreover, any microstructural change leads to changes in the distribution of residual stresses. Batista et al. (2000) suggested that the influence of residual stress is more pronounced on the propagation of the contact fatigue cracks than on their initiation. A numerical model, developed by Batista et al. (2000) for the residual stress pattern in carbonitrided gears, predicts that residual stress change starts after a few minutes of testing, the maximum level of compressive residual stresses moves towards the surface and it reaches the surface after a certain number of cycles. Since 1961, when Bush et al. (1961) brought experimental evidence, it is generally agreed that residual stress increases with testing time towards a threshold. Typical curves for residual stress pattern in gears, before and after running are shown in Figure 3-11.

A mechanism which explains the threshold has been proposed by Voskamp et al., 1980. Under cyclic loads, which do not exceed the yield stress, yielding can occur in micro-volumes where the induced stress field combined with the existing residual stress field exceed the yield limit. At the speeds involved, the load is applied for a very short period of time and during unloading the yielded micro-volumes generate residual stresses. The density of yielded micro-volumes increases with the number of cycles and so residual stresses accumulate. Theoretically, after a certain number of cycles no further yielding would occur and consequently, no further changes in the residual stress pattern. But further changes in the residual stress pattern do occur. These are explained by the microstructural changes. The transformation of retained austenite is accompanied by a volume expansion, which generates compressive stresses in the region of decomposition. The martensite decay is associated with a volume contraction, which generates tensile stresses in the region of decay. The resulting residual stresses depend on the volume fractions of the transformed phases. Based on the observed decrease in microhardness, Voskamp et al (1980) suggested that a decrease in yield stress also occur. Consequently, the density of yielded micro-volumes increases resulting in an increase in residual stress peak value and a broadening of the residual stress pattern. On the other hand, the residual stress peak value decreases due to lowering the yield stress. The final residual stress pattern thus depends on the dominance of one or other of these effects.

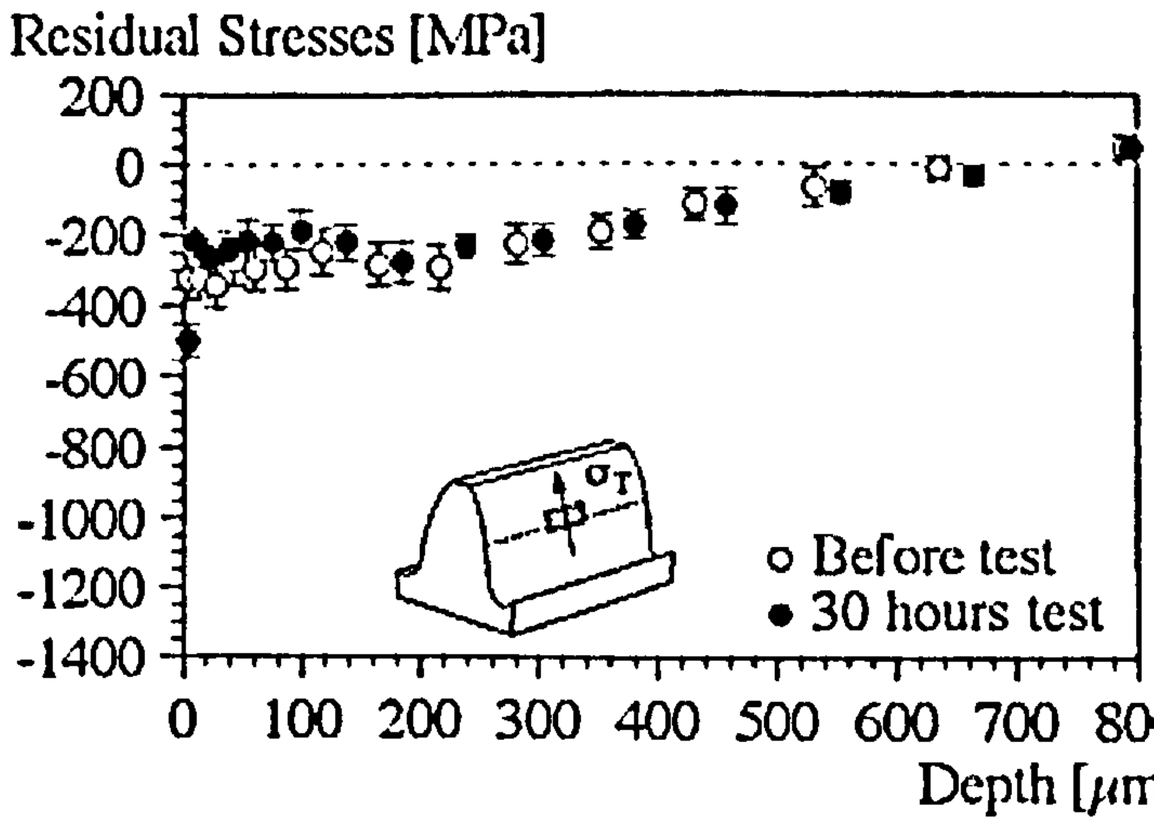


Figure 3-11. Residual stress, before and after fatigue test in a carbonitrided gear (from Batista et al, 2000).

3.4.2 Microstructural transformations

Stress and temperature conditions are favourable to atom diffusion and dislocation motion, which make possible the microstructural transformations. Whether these transformations are stress induced or they are dictated by the heating caused by friction is a debatable matter but, since both may induce transformations, it is likely that it is a combined effect. It is trivially known that under the applied stress retained austenite transforms into martensite. The stress field induced in gears by contact stresses and perhaps the heat generated by friction lead to a gradual transformation of retained austenite into martensite. This is a non-tempered martensite and consequently is very brittle. A phenomenon of martensite decay leading to the formation of specific products known as dark etching regions and butterflies has been also reported in the literature. The retained austenite transformation and the decay of martensite are treated in the next sections.

3.4.2.1 The transformation of retained austenite

The retained austenite in carburised steels is thermodynamically unstable and, under the influence of the cyclic stresses during the service operation of the gears, transforms into martensite. Krauss (1995c) distinguishes between two different mechanisms of deformation-induced transformation depending on the amount of retained austenite present in the case of carburised gears. At low amounts, the transformation does not require new nucleation sites as the stress increases towards the elastic limits at low strains. The transformation is called *stress-assisted*. At larger amounts of retained austenite larger amounts of strain are required to generate nucleation sites for the deformation-assisted transformation. Thus, an incubation strain precedes the martensite formation. This type of transformation is referred to *as strain-assisted*. In quenched and tempered carburised gear steels with a typical volume fraction of 20-30 % of retained austenite the mechanism of retained austenite transformation is stress-assisted, as Kraus (1995c) suggests.

3.4.2.2 Martensite decay

In elements subjected to cyclic surface loading (i.e., rolling bearings) specific microstructural transformations have been reported to occur. They are recognisable due to different etching characteristics when specimens are observed in light microscopy. There is an abundant literature referring to these transformations in bearings but no such transformations have been reported in gears. Since martensite decay is most often associated with micropitting or other surface contact fatigue failure, although, according to Bush et al., (1961) it can develop without its occurrence, it is tempting to believe that similar transformation takes place in gears. The generic term of martensite decay is used to describe the transformation of martensite into a number of new phases.

The region where these transformations occur is related to the depth of maximum orthogonal shear stress (Voskamp et al., 1980;) or maximum unidirectional shear stress (Martin et al., 1966; Voskamp, 1985) on one hand, and to the stress raisers (i.e., carbides and non-metallic inclusions) on the other hand. In this work the products of martensite decay correlated with the depth of maximum shear stress will be referred as: *dark etching regions* (DER) and *white etching bands* (WEB), and the products corresponding to stress raisers as *butterflies*.

3.4.2.2.1 Dark etching regions

Dark Etching Region (DER) represents a mixture of residual martensite with a ferritic phase that contains carbon in excess (corresponding to that in martensite) which is inhomogeneously distributed (Swahn et al., 1976). Martensite decay was first described by Jones (1946), who observed a dark etching region beneath the ball track of bearing inner races and claimed the occurrence of pitting as a consequence. He also claimed that this microstructural transformation was a tempering phenomenon caused by the heat generated by elastic hysteresis as the subsurface material was cyclicly stressed. The microstructural transformation is observed by a more rapid etching attack and darkening of the transformed material (see Figure 3-12). At higher magnification, the microstructure has the appearance of tempered martensite (so called by Bush et al., 1961), *troostite* (by Jones, 1946), or ferrite (by Martin et al., 1966). This type of microconstituent is only revealed by a nital etch (Österlund & Vingsbo, 1980).

Gentile et al. (1965) claimed that the DER is caused by stress concentration at non-metallic inclusions. O'Brien & King (1966b) have made a crucial observation that dark etching regions are not common to all steels. They found DER in steels which are low in non-metallic inclusions. Thus, the Gentile et al. theory does not stand.

3.4.2.2.2 White etching bands

Highly deformed regions with the appearance of thin elongated stringers, inclined at 20 to 30° to the surface, lightly etched by nital, have been observed in light microscope (Buchwald & Heckel, 1968; Martin et., al 1966). The best resolution is achieved by etching with picral solution (Gentile et al., 1965). They were called *gray lines* by Jones and *cell ferrite* by Österlund & Vingsbo, 1980. At a later stage another set of bands, much larger than the previous ones and inclined at 70-80° to the surface have been observed by Swahn et al., 1976. By reversing the direction of rolling, the white etching bands will develop with a new orientation (Voskamp, 1985). This suggest a stress-dependent transformation.

The **30° bands** (Figure 3-12b) are disc-shaped regions of ferrite inclined at 30° to the surface bordered by carbide discs which are constituted by very small carbide particles (Swahn et al., 1976).

The **80° bands** (Figure 3-12c) are disc-shaped regions of plastically deformed ferrite inclined at 80° to the surface.

The structure of the WEB resembles a heavily cold rolled ferrite (Österlund & Vingsbo, 1980) and exhibits a cell-type structure free of carbides (Gentile et al., 1965; Martin et al., 1966). Tempered carbides are not present in the WEB (O'Brien & King, 1966a; Österlund & Vingsbo, 1980) and the amount of proeutectoid carbides is gradually reduced during cycling (Buchwald & Heckel, 1968). Muroga & Saka (1995) determined the lattice constants of the WEB polycrystals and they suggested it can be either α-Fe or martensite. They also suggested the possibility of the dissolution of cementite and/or austenitizing followed by re-formation of martensite because of the higher content of Mn and Cr found in WEB.

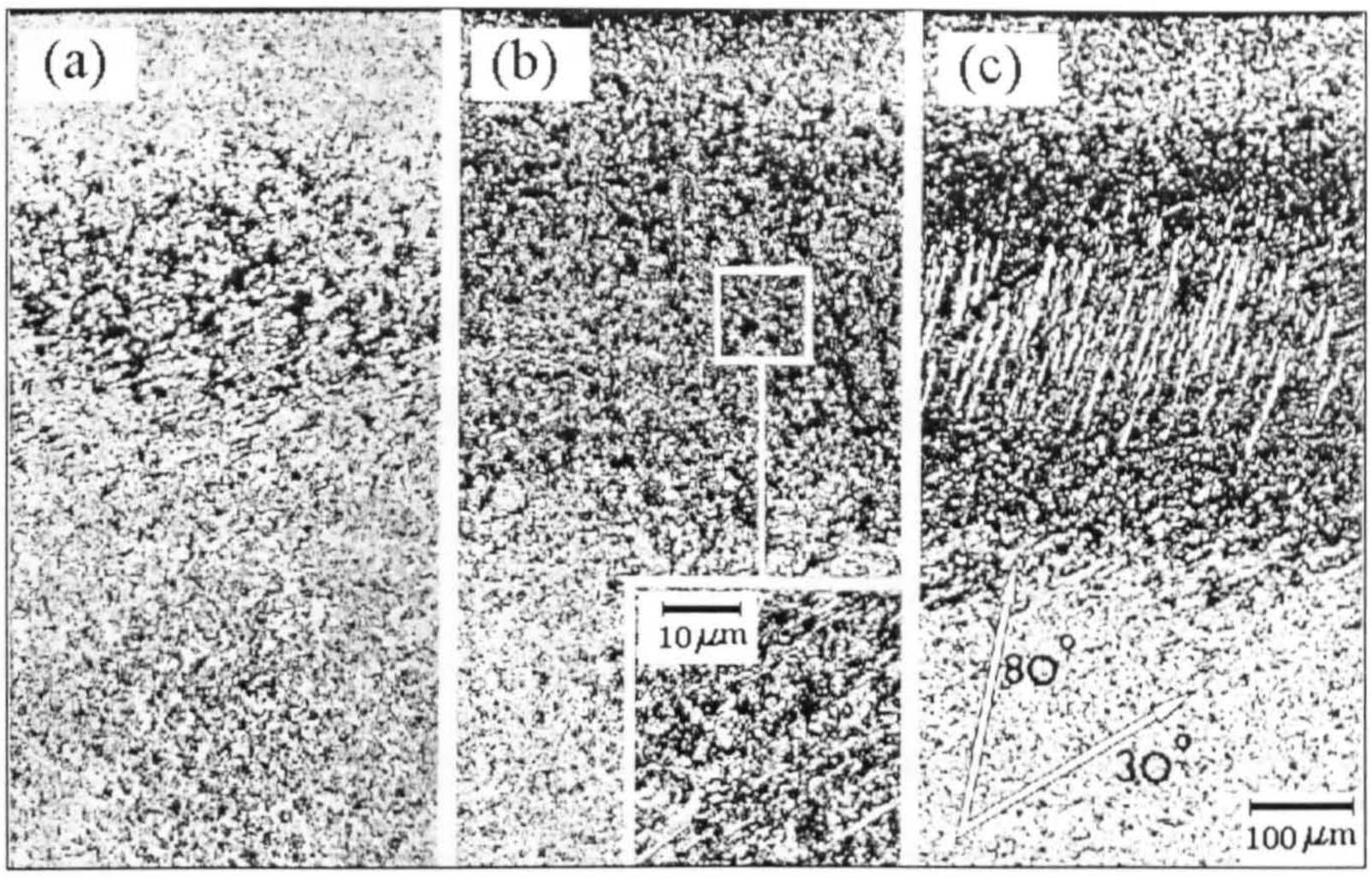


Figure 3-12. The three stages in martensite decay: (a) the formation of DER, (b) the formation of 30° bands and, (c) the formation of 80° bands (after Swahn et al., 1976).

3.4.2.2.3 The butterflies

The decay of martensite caused by stress raisers is represented by the so called butterflies, first described by Styri (1951). A butterfly is a double wing-like region, associated with non-metallic inclusions or carbides, which is light etched compared to the martensite matrix (see Figure 3-13).

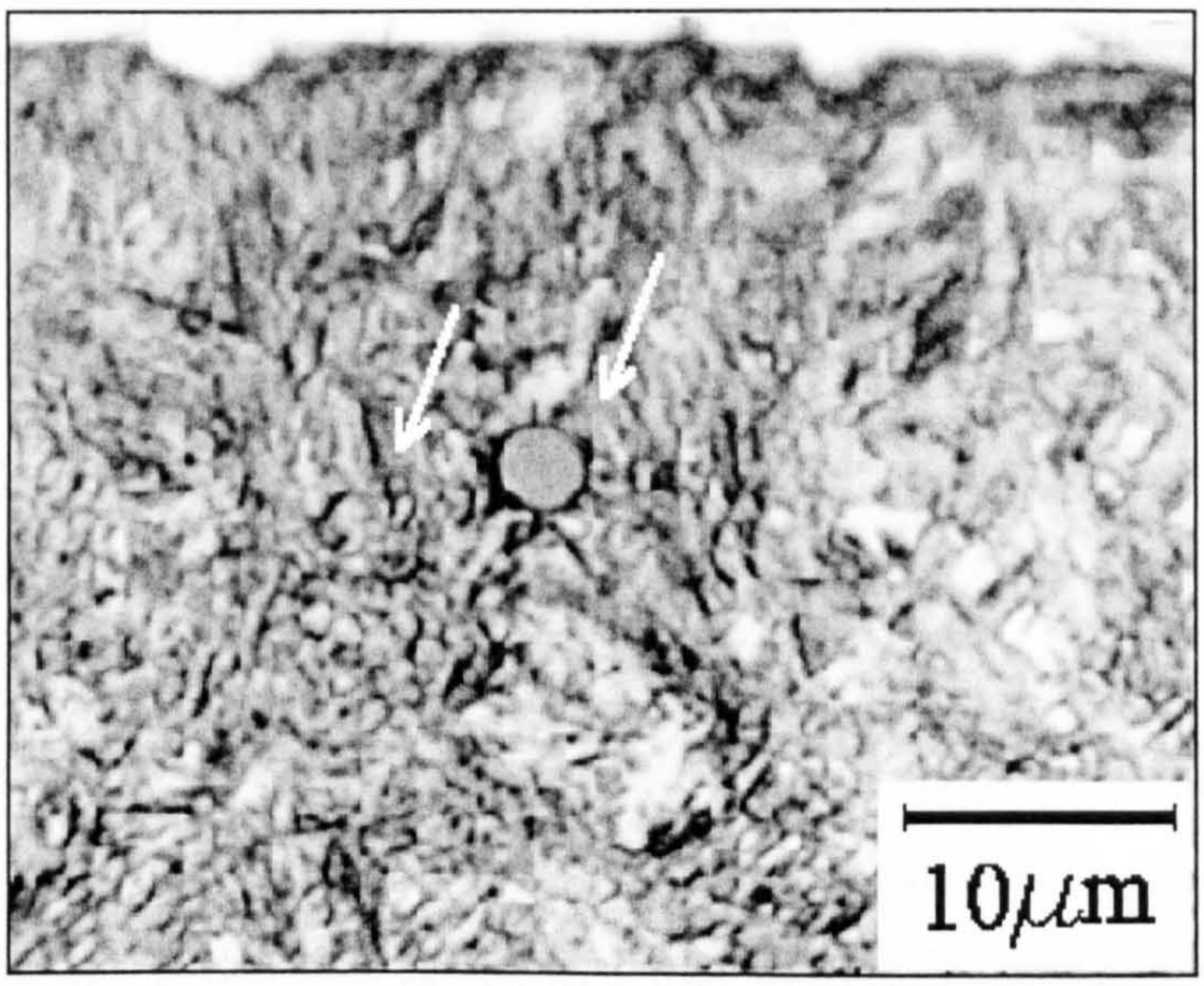


Figure 3-13. Butterfly found in a fatigued gear (16MnCr5 steel). The wings are marked with arrows.

The butterfly wings are oriented at a 40-45° angle relative to the rolling direction. Similar to the white etching bands, when the direction of rolling is reversed, a second pair of wings of symmetric orientation is developed (Österlund et al., 1982). This indicates that the transformation is stress induced. The stress concentration factor generated by an inclusion depends on the size, shape and physical and mechanical properties of the inclusion.

Butterflies are formed around oxides, silicates and nitrides of molybdenum, vanadium, iron, titanium and chromium, and also around primary carbides. Fine carbide particles and sulphide inclusions are reported not to form butterflies (Österlund et al., 1982, Littman & Winter, 1966). The butterfly wings exhibit higher microhardness than that of the martensite, possibly, because of the precipitation hardening (Österlund et al., 1982). TEM investigations of the butterfly's wings have revealed a cellular structure similar to that found in the white etching bands (O'Brien & King, 1966b). This suggests that both types of transformations are different manifestations of the same mechanism.

3.4.2.2.4 The mechanism of martensite decay

It has been suggested that the martensite decay can be initiated within the martensite needles in the regions of high density of dislocations or at the martensite/carbide interface. The martensite is a supersaturated solution of carbon in iron and is thermodynamically unstable. Due to the inhomogeneous plastic deformation in localised areas during the fatigue process, carbon atoms are rejected from martensite and they migrate towards regions with a high density of dislocations. After sufficient stress cycles, carbon diffusion can be induced. Carbon in ε-carbides and in martensite migrates to dislocations impeding their motion. This idea is supported by the observed increase in microhardness in the early stages of transformation (Swahn et al., 1976). Carbon from martensite precipitates along the coarse slip lines. Therefore, within the slip bands only highly deformed ferrite is left. The dislocation structure, studied by TEM (Voskamp et al., 1980) has confirmed that the DER-ferrite is heavily deformed. The new phase, called *DER-ferrite*, grows in groups of parallel discs of arbitrary orientation, separated by martensite. Other possible site for nucleation is the carbide-martensite interface (Bush et al., 1961).

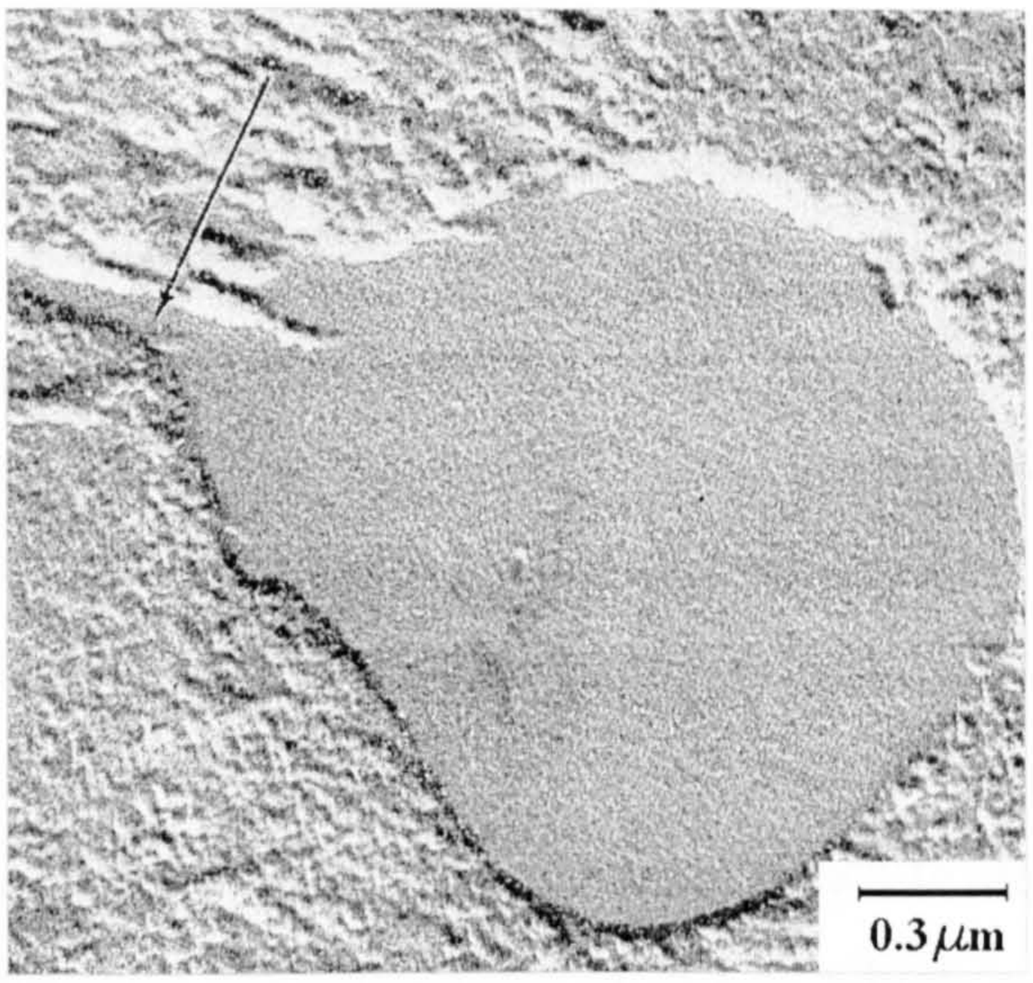


Figure 3-14. An extrusion (arrowed) at a carbide particle (after Bush et al., 1961).

Martensite needles situated in the vicinity of a carbide particle will yield as soon as the load is applied. Consequently, dislocations pile up at the carbide matrix interface where they can react and form vacancies. When enough vacancies are formed at the phase boundary the conditions for extrusion mechanism are reached. This continues until the entire carbide particle has been extruded into thin plates. Figure 3-14 shows such an extrusion in a carbide particle. The DER-ferrite is more ductile than the martensite and, consequently, during continued cycling, local plastic deformation occurs. Dislocation slip gives rise to shear along a well defined direction, given by the stress tensor according to Swahn et al., (1976) or by the crystallographic directions according to Voskamp, (1985). Each strain nucleation site forms a disc-shaped deformation zone, inclined at 30° to the rolling direction. Carbon solubility is increased in the deformed zone due to the presence of dislocations and the localised deformation process allows the migration of carbon away from the deformed region, leading to the formation of *lenticular carbides* (Buchwald & Heckel, 1968). These carbides serve as precipitation site for carbon from the DER-ferrite. Thus, at the interface of the deformation zone, thin carbide discs are being formed. These regions of deformed ferrite bordered by thin carbide discs are called white etching bands (WEB).

No satisfactory mechanism has been yet developed. Swahn et al. (1976) suggetsed that, as cycling continues, the 30° bands grow and the distance between them becomes

smaller and smaller. It is possible that, when the space between the 30° bands is sufficiently small, the carbide discs break-up and the nucleation of 80° bands starts.

In addition to the above theories, a mechanism based on a texture development has been proposed by Voskamp (1985). He found that the main preferred orientations coincide with a (100) [110] texture relative to the rolling direction. The [110] direction of BCC-Fe becomes aligned parallel to the surface and to the rolling direction while (100) planes are aligned perpendicular to the surface and at +45 and -45° to the rolling direction. This development is noticeable before the occurrence of white etching bands. The angle of orientation coincides with the orientation of WEB, which suggest (on contrary to Swahn et al. theory) that the bands have a crystallographically determined nature rather than a stress tensor correlated nature. The local change in texture determines the direction of crack propagation. This idea is supported by the fact that the (100) plane is a common fracture plane in the steel matrix (Voskamp, 1985). The driving force for crack propagation is given by the residual tensile stress, which acts during unloading. A finite element model created and used by Bhargava et al. (1990) to calculate the inclinations and amount of plastic cyclic strain showed a good correlation between the inclination of the white etching bands and the inclinations of concentrated plastic activity.

A point of confusion relates to the microhardness of the transformed region. It was reported to be higher than that of the matrix by some authors (Choi & Bahng, 1998) while others reported to be lower (Bush et al., 1961; Voskamp, 1985). It appears that the conventional microhardness is not the appropriate test for such fine and inhomogeneous microstructure and only nanohardness can give reliable results.

Sometimes the white-etching bands were confused with the white-etching wings of the butterflies. The mechanism of their formation is also stress induced. Österlund et al.(1982) suggested that the transformation of butterflies takes place by a recrystallisation process. Butterflies are formed when plastic deformation is constrained by the surrounding material, and is more common when shear stresses are very high. The combination of Hertzian stresses and stresses generated by the presence of an

inclusion generates strong dislocation densities at the interface. Österlund claims that these dislocations are responsible for recrystallisation. Due to the difference of the elastic and plastic properties (i.e., Young's modulus and yield stress) between matrix and inclusion high plastic strains are induced. These strains allow dissolution of tempered carbides, therefore, nucleation of ferrite is favoured. These nuclei grow and form nanograins (see Figure 3-15a), which become stress concentration sites. With continued cycling, the dislocation density increases and, when the density is high enough the transformation starts (see Figure 3-15b). The newly nucleated grains undergo plastic deformation and work harden. In this way the stress field is redistributed and new nuclei are added in the front boundary. New nuclei are created up to the moment when the stress field level becomes so low that it can no longer be responsible for new nuclei production. The butterfly's wings may grow until they reach a neighbouring grain.

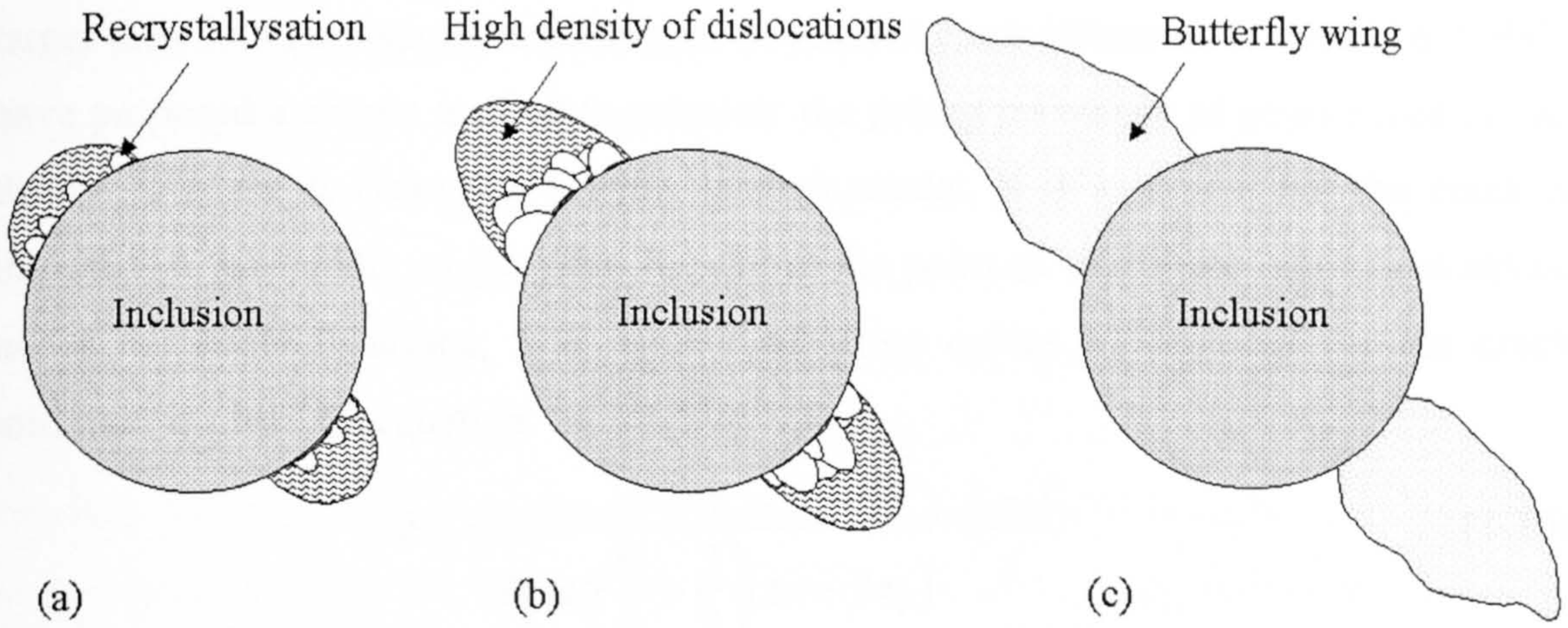


Figure 3-15. The mechanism of butterfly formation. (a) The nucleation of the recrystallysation process at the interface martensite/inclusion; (b) The growth of initial nuclei and the formation of new nuclei; (c) A developed butterfly.

3.4.2.3 Cracks in rolling/sliding contacts

In rolling/sliding contacts the direction of the principal stress and strain tensor at a point vary continuously as the contact area passes over that point. Crack initiation models that are based on uniaxial parameters are not adequate for these multiaxial conditions. A crack, under rolling/sliding conditions, is likely to initiate on "critical planes", which can explain why fatigue cracks are inclined to the surface (Ringsberg, 2001). The

models proposed for contact fatigue crack initiation are classified by Tallian (1992) in two groups: *engineering models* and *research models*. In engineering models the contact material is considered as a matrix with defects and fracture mechanics is used to calculate the number of cycles required for a crack to propagate to a critical size. The cyclic stresses applied by the contact produces irreversible changes near the defects and, at the most severe defect a micropit will form. In research models, the material is considered free of defects and aim to estimate the total fatigue life. A review of the models outlined above is beyond the interest of this work but the reader is advised to see the work of Tallian (1992). As already mentioned, cracks can initiate at the surface or under the surface. We adopt the terminology used by Glodez et al (1997) to define surface and subsurface cracks. Cracks initiating at a depth ranging from 0 to 2 μm are classified as surface cracks. Cracks initiating at a depth from 2 μm to 0.4*b* (*b* is the half-width of the Hertzian contact) are called subsurface cracks. Cracks initiating at depths larger than 0.4*b* correspond to the regime of internal crack initiation. Glodez et al (1997) have proposed a simple method to calculate the pitting resistance of gears based on the theory of Mura & Nakasone (1990). In their model, it is assumed that the crack is initiated in the middle of the grain located at the point of maximum equivalent stress, under the contact surface. The number of stress cycles N_0 required for the crack initiation is given by equation

$$N_0 = \frac{\gamma}{h\left(\log\frac{8l}{h} - \frac{3}{2}\right)(\Delta\tau - 2\tau_f)} \cdot \frac{2-f}{f} \tag{3-10}$$

where, γ is the surface energy of the crack, h is the width of a slip band, $2l$ is the length of dislocation pile-up (and can be approximated to the grain diameter), $\Delta\tau$ is the applied stress amplitude on the slip layer, τ_f is the frictional stress of the material and, f is the irreversibility factor of dislocation pile-up.

3.5 Micropitting characteristics

Micropitting occurs under conditions of partial elastohydrodynamic lubrication where a significant asperity contact takes place. It is most often associated with case hardened spur gears but all gears are susceptible, including helical and bevel. Micropitting can occur with all heat treatments applied to gears (Errichello, 2002): carburising, induction hardening, through hardening and nitriding. Although the nature of the lubricant influences micropitting, the problem appears with both synthetic and mineral lubricants (Shotter, 1980). The first indication that micropitting has occurred is the increasing noise and vibrations made by the operating gears due to the deviations of the tooth profile. The severity of this type of damage implies that the micropitting could represent the initial stage of any other type of gear tooth failure. Shotter (1980) emphasised that, "All of the classical gear tooth failure modes can sometimes have their origin traced back to micropitting." The above considerations explain the great current interest in micropitting.

A micropitted surface appears dull to the naked eye, as if it has been etched. Because of this, the first names given to this type of failure were *frosting* and *grey staining*. Micropitting is also referred to as: *glazing*, *microspalling*, *peeling* and *superficial spalling*. Microscopically, micropitting shows similar features (i.e. pits and cracks) as macropitting but these differ in scale. It occurs most often on the dedendum section of spur and helical gears, but it is very often reported on both the dedendum and addendum. It may even extend along the pitch line (Brechot et al., 2000). The two aspects of micropitting phenomenon that can be observed on the gear tooth surface: *micropits* and *microcracks* are shown in Figure 3-16. A micropit is characterised by a depth, length and width on the order of few microns or tens of microns. The microcracks have different orientations in dedendum and addendum, which suggests that their orientation depends on the sliding direction. On the driving gear (pinion) the cracks propagate towards the pitch line meanwhile on the driven gear (wheel) cracks propagate away from the pitch line. Unlike in classical rolling fatigue, in micropitting, the crack direction is independent of the contact motion, depending only on the direction of the friction force (Olver, 1983).

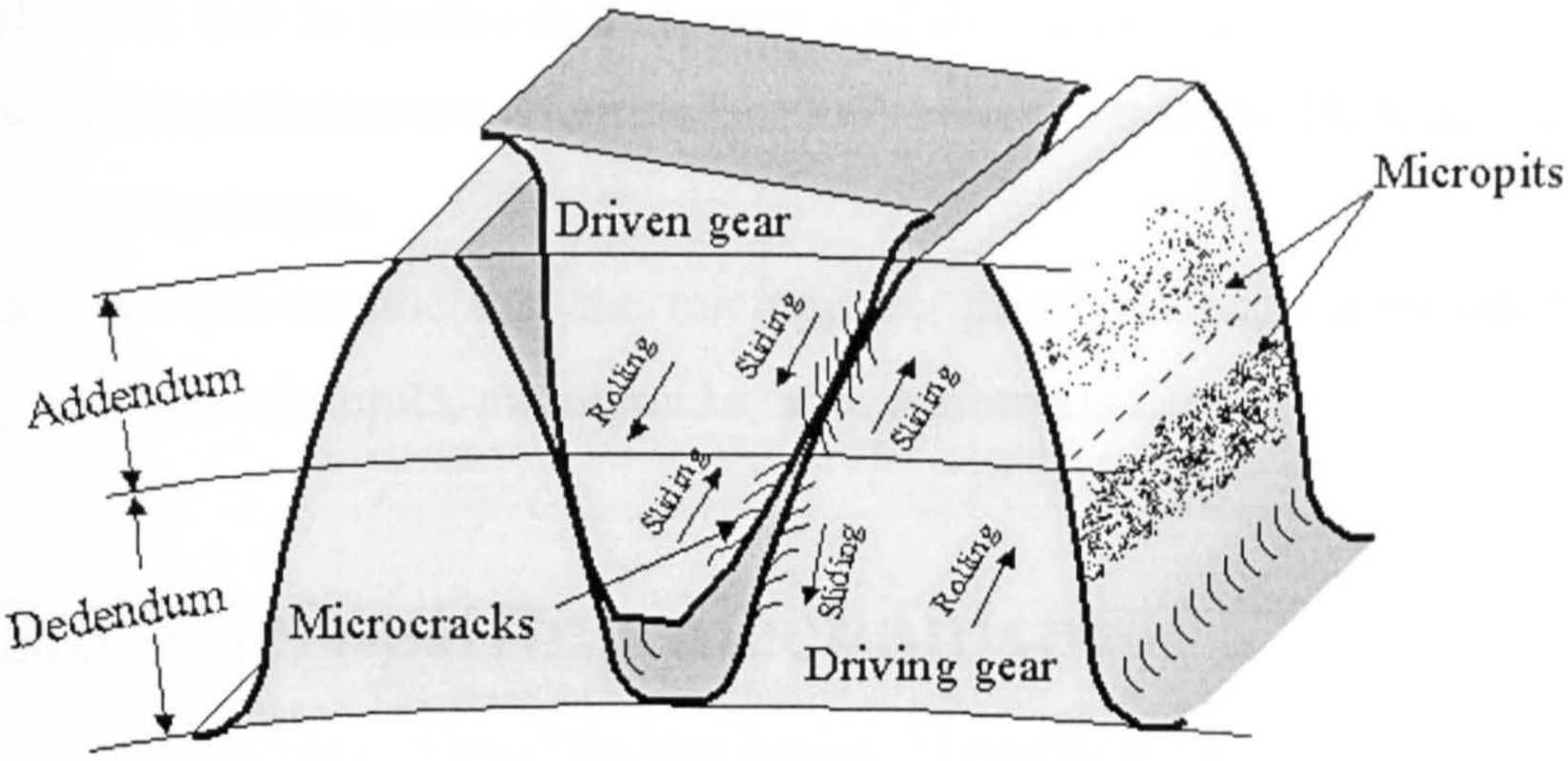


Figure 3-16. The orientation of microcracks changes at the pitch line according to the sliding direction.

These cracks propagate into the depth of the steel at a shallow angle, usually less than 45°. In pure rolling discs tested by Berthe et al. (1978) they found micropits having a depth of 20 μm and macropits having a depth of 200 μm. They concluded that the depth of macropits corresponds to a characteristic Hertzian depth, meanwhile the micropits resulted from asperity contact. Their justification is based on the fact that micropitting occurs for specific film thickness higher than 1, i.e. $\lambda > 1$. The location of microcracks according to Zhou et al. (1989) is less than 10-20 μm below the surface and the depth of micropits ("the depth of failure") is about 30-60 μm below the surface. This depth is much closer to the surface as compared to the maximum shear stress, which is typically at 250-300 μm depth. Tests performed by Nelias et al. (1999) on a disc machine enabled them to determine certain conditions for micropitting occurrence. By the means of ultrasonic echography they found that many microcracks were nucleated after $5 \cdot 10^5$ cycles, and butterflies and microcracks appeared around primary carbides and non-metallic inclusions (alumina, calcium aluminate). The depth of maximum concentration of butterflies was about 200-300 μm, much smaller than the depth of maximum Hertzian shear stress, which was 457 μm in their study.

Micropitting in rolling element bearings is almost always associated with microstructural transformations. To the author's knowledge, the dark etching regions in gears have been only reported by Hoeprich (2000). However, he points out that the dark

etching effect is due to plastic deformation and dislocation accumulation rather than microstructural transformation. Hoeprich (2000) believes that the DER are evidence of the presence of hydrogen.

In summary we can assume that the micropitting phenomenon is associated with the following features: micropits, microcracks, and, martensite decay.

3.6 Micropitting mechanism

The first micropits occur on the tooth surface at the peaks of the surface roughness, consequently, the waviness becomes smaller. Because of the effect of surface finish improvement micropitting was described in the past as a beneficial phenomenon. Since the pitch line is rarely affected by micropitting it appears that sliding is necessary for the damage to occur in gears. However, micropitting occurs under pure rolling conditions, in rolling element bearings. Figure 3-17 shows the sequence of events in the development of a micropit. A crack is initiated at the surface or below the surface (1), it grows for few microns on the planes that are inclined to the surface at angles between 5 and 50°. After this growth the crack may branch (2) with one branch propagating parallel to the surface (2→3) and the other one (2→5) continuing to propagate into the bulk. This may become the origin for a macrocrack, which may initiate the classical pitting. The branch parallel to the surface grows for some distance before it turns back to the surface. At the point where the crack reaches the surface (4) a micropit is being formed.

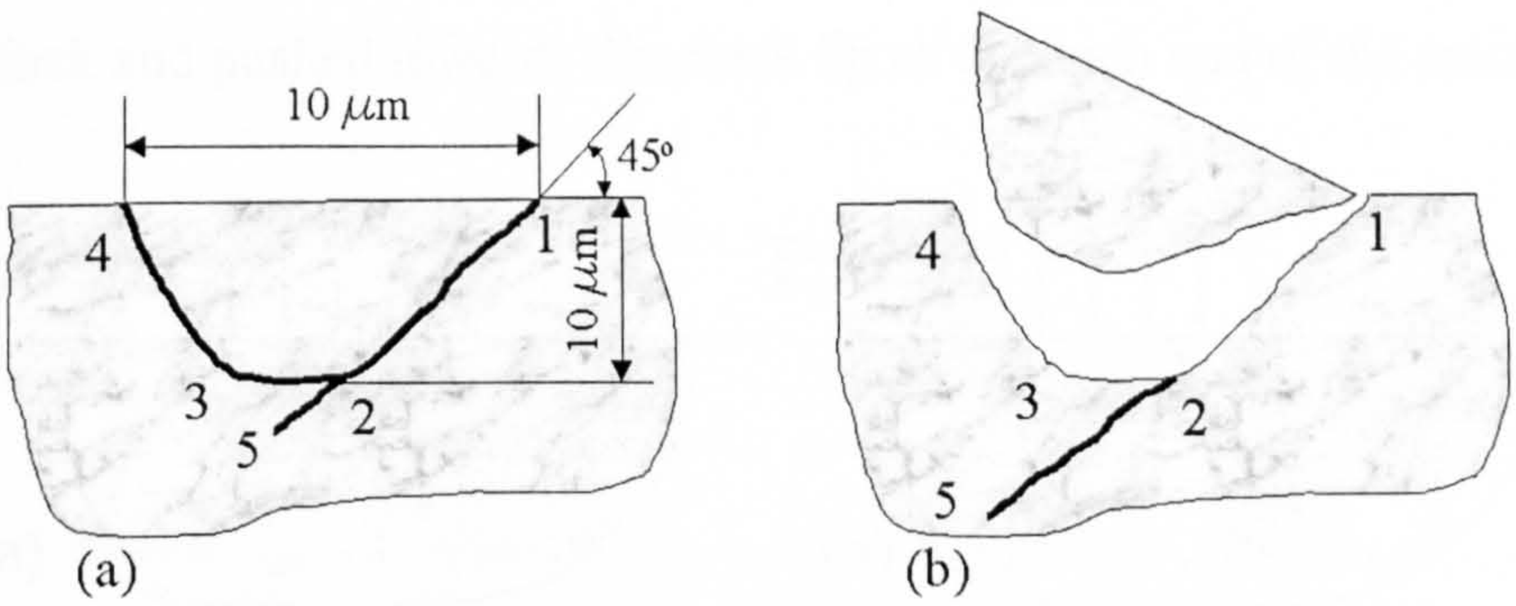


Figure 3-17. Schematic representation of micropitting: (a) microcrack propagation, (b) micropit formation.

Micropitting is thus a phenomenon driven by crack initiation and propagation. Since the micropitting morphology resembles macropitting one could assume that the mechanisms are also similar. The pitting mechanism was described for the first time by Way (1935). He measured the number of cycles for a pit to occur on the surface of rollers in lubricated contact. Based on the fact that no pitting was seen to occur in dry contacts, Way launched the theory of "oil penetration". According to this theory the pits are formed by the propagation of surface originating cracks under the influence of contact stress and pressurization of the crack by trapped lubricant. Preferential sites for surface microcrack initiation are represented by surface irregularities such as grinding marks. The microcracks are caused by high local shear stress at the surface roughness peaks and high local friction forces due to mixed lubrication. As shown in section 2.5.10, about 10 % of the total area of contact is represented by asperities. This leads to values of friction coefficient as high as 0.06 while smooth surfaces give 0.015. The high values of friction coefficient impose important tangential stresses on the asperities, therefore, fatigue failure can be initiated. The lubricant viscosity affects the amount of fluid that flows into the crack. A lubricant with a sufficient high viscosity can prevent pitting since the lubricant cannot penetrate the cracks. The presence of the lubricant inside the crack has been confirmed (Bower, 1988) by secondary-ion mass spectroscopy. The author cited above proposed three possible mechanisms of crack propagation: the shear mode, in which crack propagation is driven by the cyclic shear stress, the "hydraulic pressure mechanism" (see Figure 3-18a), in which fluid is forced into the crack by the load and the hydraulic pressure pushes apart the faces of the crack and, the "fluid entrapment mechanism" (see Figure 3-18b), in which the fluid is trapped inside the crack and pushed towards the crack tip as the open end of the crack is closed.

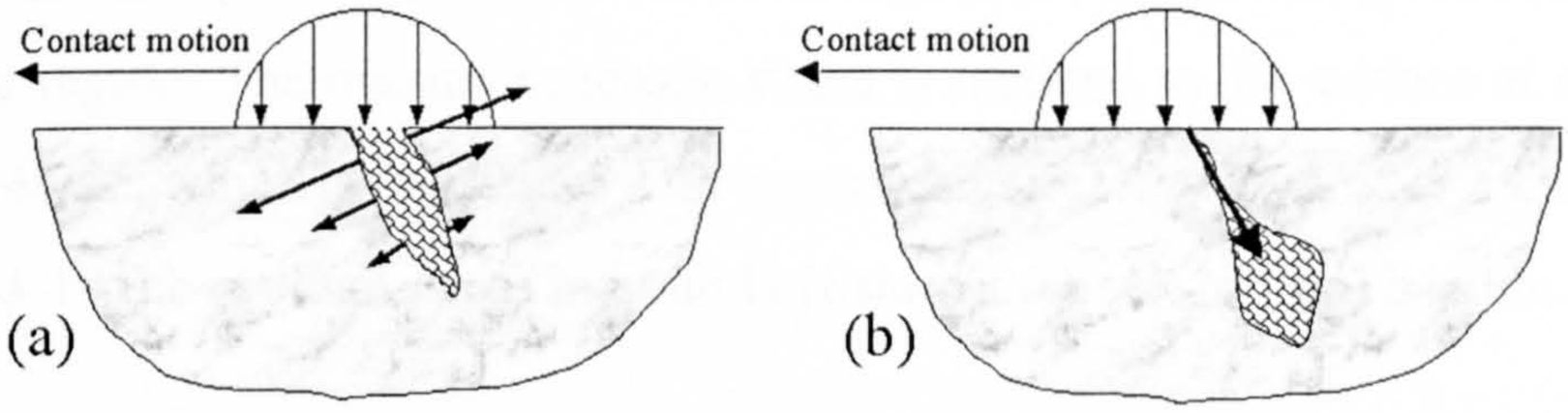


Figure 3-18. Mechanisms of crack propagation: (a) the "hydraulic pressure" mechanism and (b) the "fluid entrapment" mechanism.

The hydraulic pressure theory does not explain the propagation direction of the cracks (i.e., the direction opposite to that of the friction force).

Berthe et al. (1980) have proposed a micropitting mechanism under pure rolling conditions, which involves the local elastic stress field due to mutual asperity contact and a criterion based on the normal elastic asperity stress. The asperity behaviour (i.e., elastic, plastic or, elasto-plastic) is determined by the variation of the dimensionless roughness ratio, h_r/σ, versus the plasticity index, Ψ, defined as

$$\Psi = \frac{E'}{p}\sqrt{\frac{\sigma}{\beta^*}} \tag{3-11}$$

where, h_r is the distance between the mean lines of the two surfaces before deformation of asperities, σ is the composite roughness, p is the asperity pressure and β^* is the mean radius of asperity curvature.

The results showed that when the asperity pressure decreases to a value in the range of elastic deformation, the rate of micropitting tends to zero but no correlation was found between asperity plastic deformation and micropitting. However, this model does not explain the dependence of crack direction on the sliding direction (i.e., the direction of tangential force). The direction of the crack is such that the friction force acts to close rather than open the crack. Olver (1983) suggested that the crack formation in micropitting is a consequence of the plastic deformation of surface asperities. The combined effect of normal and tangential stress leads to a compressive yielding of the asperity, which results in a tensile residual stress. Evidence of plastic deformation of asperities can be found in different etching characteristics (i.e., lighter than the matrix, according to Olver) that occur in the vicinity of a crack. The model is based on Green's slip-line field and predicts a region of tensile residual stress, during unloading, in the subsurface region. The maximum tensile stress is inclined to the surface at an angle in the range 4-30°, which is consistent with experimental observation of crack orientation. In Figure 3-19 the residual stress in point D is shown for $\varepsilon = 20°$ and friction coefficient $\mu = 0.2$.

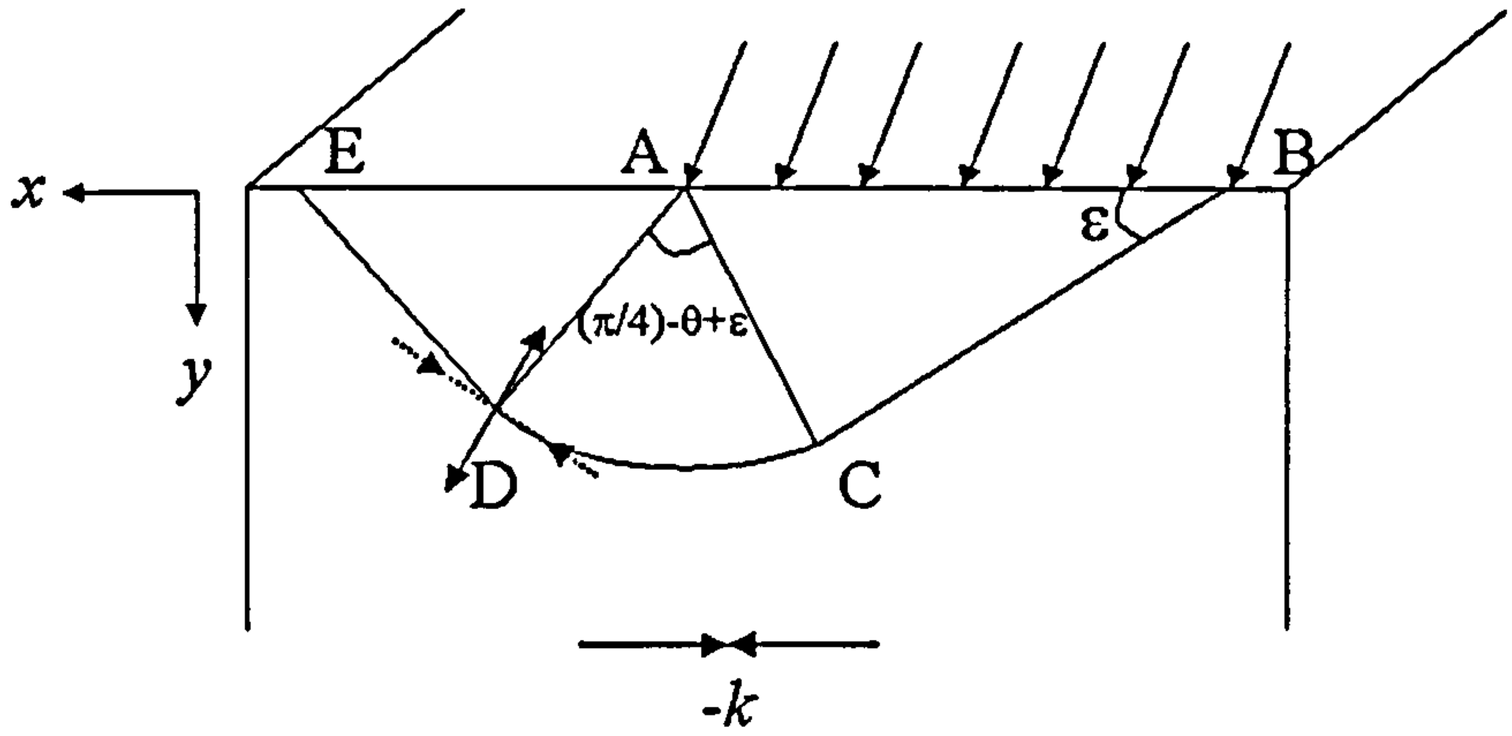


Figure 3-19. Slip line field for $\varepsilon = 20°$ and $\mu = 0.2$. k is the yield stress in pure shear. Tensile stresses are drawn in solid arrows.

The friction coefficient is given by

$$\mu = \frac{\tau_{xy}}{\sigma_y} = \frac{\cos 2\varepsilon}{1 + (\pi/2) - 2\theta + 2\varepsilon + \sin 2\varepsilon} \tag{3-12}$$

where, τ_{xy} is the tangential stress on AB, σ_y is the normal stress on AB, θ is the slope of the undeformed asperity.

The concentrations of local stresses and the plastic deformation around sub-surface or surface inhomogeneties lead to a local accumulation of dislocations, which move along the slip bands. Three possible sites of surface crack initiation are described by Zhou et al. (1989). Microcracks may initiate along slip bands by dislocation motion, along grain boundaries (carbide particles or retained austenite), or along the interfaces of inclusions. The authors reported the formation of a highly deformed layer very close to the surface. They found a high dislocation density in tempered martensite located below the surface, which extends 0.2 to 0.7 μm from the contact surface. They concluded that fatigue cracks initiate at the intersection of slip bands, grain boundaries, inclusions or other interfaces with the free surface. Chue et al. (2001) have described the formation of three distinct layers during rolling/sliding contact. They are the *ratchetting layer*, the *elastic/plastic shakedown layer* and the *elastic layer* in order starting from the surface. The ratchetting layer (term introduced by Kapoor & Johnson, 1994) is about 5 μm and is characterised by a steady stress state of open elastic-plastic loops. The elastic/plastic shakedown layer is about 5 μm and is formed due to the closed elastic-plastic loop. These two layers form the *work* or *strain hardened layer*. The authors correlate the

thickness of this layer with the hydraulic pressure for creating a pit and conclude that the required hydraulic pressure decreases as the thickness increases.

Sub-surface microcracks are initiated due to repeated traversing of the contact load, and stress concentration around inhomogeneties (non-metallic inclusions and carbides). The stress concentration is explained by the elasto-plastic strain mismatch between inhomogeneties and the martensitic matrix. When the stress reaches the yield point of the matrix a plastic strain develops in a small region surrounding the inhomogeneity. Due to the repeated contact, dislocations move back and forth and accumulate in this region. Localised microstructural changes (e.g., butterflies) in steel may then occur and a crack is nucleated. The initial size of microcracks is determined by the size and nature of inclusions. This initial microcrack may be attracted or repelled by the closest inclusion and therefore starts to propagate.

3.7 Factors influencing micropitting

Extensive research has been done in order to identify and monitor the factors that influence micropitting initiation and propagation. In this work it has been found more convenient to group these factors into two classes: *materials related factors* and *operating conditions*.

3.7.1 Materials related factors

This category of factors refers to the chemical composition and microstructure of the gear steel and the lubricant chemistry.

3.7.1.1 Chemical composition and microstructure of the gear steel

Winter and Weiss (1981) have shown that gears with case carbon content of 0.6 % or less are especially prone to micropitting and micropitting resistance increases with the case depth to a depth limit beyond which no further increase occurs. The chemical composition of the steel influences the transformations that take place during heat treatments, hence the final microstructure. The microstructural constituents in the case

of a gear steel are: martensite, retained austenite, carbides and non-metallic inclusions. The effect of retained austenite and carbides is debatable while the effect of non-metallic inclusion is known to be harmful.

3.7.1.1.1 Tempered martensite

Many believe that tempered martensite gives optimum fatigue properties. The higher the carbon content the higher is the density of the transition carbide clusters. These clusters, with the associated dislocations increase the hardness and strength of martensite. Olver (2001) reported that even small hardness differences between the two surfaces have an important effect. Since gears are, usually, nominally identical, small hardness differences may explain the apparent random distribution of micropitting. Tests carried out on discs with different hardness have revealed that the softer one is more prone to micropitting (Olver, 2001).

3.7.1.1.2 Retained austenite

Some investigators believe that retained austenite, in moderate amounts, is beneficial to fatigue resistance, while others believe that retained austenite should be minimised for optimum fatigue performance. Typically, the retained austenite content in case microstructure of direct carburised steels is between 20 and 30 %. Too much retained austenite (~40 vol. % or more) is detrimental to the performance of carburised steels. A high content of retained austenite may reduce surface compressive residual stresses and contribute to a decrease in fatigue resistance. Krauss (1995b) believes that dominant sites of fatigue crack initiation are prior austenite grain boundaries. Winter and Weiss (1981) support the idea that retained austenite in the surface layers has a beneficial influence on micropitting resistance. Szpunar & Bielanik (1984) published a review of the work done on the influence of retained austenite. They concluded that an amount of 30-50 % of retained austenite in the carburised cases of gear tooth has favourable effect on the contact fatigue. This influence is explained by the increased ability of the steel to undergo plastic deformation. Their research also revealed that the maximum crack propagation rate is achieved in specimens containing 23 % of retained austenite. Retained austenite gradually transforms into martensite during contact. The associated

work hardening reduces the rate of crack propagation. In addition, the transformation of retained austenite is accompanied by a volume expansion, which generates compressive stresses which are also beneficial for the fatigue performance of the gear steel.

3.7.1.1.3 Carbides

Zum Gahr (1987) reported that the wear loss increases with an increased volume fraction of carbides. Not only the carbide content contributes to increasing wear loss but also the carbide shape and size are very important. In order to increase fatigue resistance the carbides should be very small and of spherical shape. The combination of a high content of retained austenite and coarse carbide particles leads to reduced resistance to fatigue crack initiation. As discussed in the mechanism of martensite decay section, the carbide interface may represent the initiation site of the transformation.

3.7.1.1.4 Non-metallic inclusions

It is well established that non-metallic inclusions can play an important role in initiating cracks. However, to assess their role in contact fatigue is complicated due to their different nature, size, morphology and distribution. Non-metallic inclusions act as stress raisers. When localised plastic deformations occur around an inclusion the nucleation of a crack becomes possible. Non-metallic inclusions are at the origin of sub-surface micro-cracks at the Hertzian depth, independently of the level of surface roughness (Nelias et al., 1999). The most detrimental types of non-metallic inclusions are oxides. A higher density of inclusions (e.g. carbides) can be detrimental due to the possibility of crack nucleation by the interaction of several inclusions. As already discussed, the formation of butterflies is closely related to inclusions. Microcracks are very often associated with butterflies. The hard inclusions are much more detrimental to fatigue life than the soft ones. Cracks can originate at inclusions but the rate of crack propagation is a function of the matrix properties.

3.7.1.2 Surface roughness

Simple experimental observations show that the rougher is a surface the more prone is to micropitting. Surface asperities act as stress raisers and surface initiated cracks originate in the asperities. A stress bridge may be created by the superposition of the Hertzian stress and localised stress due to roughness of large wavelength and a subsurface originated crack can propagate. Ariura et al. (1983) investigated micropitting by means of scanning electron microscopy (SEM). They concluded that micropitting might be avoided by a very good surface finish (no micropitting was observed on gears polished with a buffing wheel). Britton et al. (2000) compared the frictional losses of case carburised gears made from En36 steel, having different surface finish. They found a reduction of friction coefficient of about 30 % for super-finished gears (R_a~0.05μm) compared to those for conventionally ground teeth (R_a~0.4μm) with correspondingly lower tooth surface temperatures under the same conditions of load and speed. In a study using finite element modelling Batista et al. (2000) found that surface roughness influences the stress only in the first 50 μm below the gear tooth surface. This depth is much smaller than the depth of maximum shear stress (on the order of hundreds of microns) but corresponds to the depth of micropitting.

3.7.1.3 Lubricant chemistry

Micropitting can occur with mineral or synthetic lubricants. Tests conducted by Höglund (1998) allow him to conclude that the esters, polyglycol, polyalphaolephin and rapeseed oils are the best to use to avoid pitting fatigue. Lubricants with a higher viscosity reduce micropitting due to higher film thickness. Naphtenic mineral oils have compact molecules that generate high friction, whereas paraffinic mineral oils and polyaphaolefin synthetic oils have elastic molecules and low friction coefficients (Errichello, 2003). Shotter (1980) suggested that lubricant additives may have a significant influence on micropitting by modifying the friction force at the asperity scale. Brechot et al. (2000) emphasised that the use of synthetic oils such as polyalphaolefins and polyglycols can reduce the heat generated in the gear contact. The lamda ratio has little effect on micropitting initiation but has a strong influence on

micropitting propagation. Olver (1995) found that, at contact stresses of 1.7 GPa and lambda ratio $\lambda \geq 2$, no micropitting occurs.

Additives contained in lubricants may affect lubricant fluidity, crack penetration and then hydraulic pressure propagation. Brechot et al. (2000) noted that antiwear (AW) and extreme pressure (EP) additives used to prevent scuffing might promote fatigue, including micropitting. Additives containing lead or those based on sulphur and phosphorus (Winter & Oster, 1990) have been proven to prevent micropitting, whereas additives based on zinc-ditiophosphates tend to support micropitting (Höhn et al, 1996). On the contrary, Cardis & Webster (2000) have reported that nitrogen and phosphorus anti-wear additives showed a significant tendency to produce micropitting. These discrepancies can be attributed to different operating conditions used by the cited researchers. The effect of additives might change from a beneficial to a harmful one depending on the temperature, pressure, speed, etc. In a paper by Nixon (1998) referring to the effects of EP additives on the fatigue life the author concluded that these effects could override the benefits of enhanced steel cleanliness and surface finish. Additives also can cause corrosion at elevated temperature. All lubricants are susceptible to water contamination, particularly mineral oils with EP additives. Experiments have shown that wet lubricants promote micropitting (Errichello, 2003).

The ideal gear lubricant thus has to have its viscosity-temperature behaviour optimised along with its additive system (Bala et al, 2000). An ideal gear lubricant would have high viscosity, high pressure-viscosity coefficient and low friction coefficient. Unfortunately, lubricants with high pressure-viscosity coefficient tend to have a high friction coefficient.

3.7.2 Operating conditions

The current trends in gear design are towards increased efficiency, which means more severe operating conditions: higher load, higher speed and, as a result, higher operating temperature. The contact stress, operating speed and temperature, and sliding have been proven to strongly influence micropitting initiation and propagation. An effect associated with surface contact fatigue is the formation of *wear debris*.

3.7.2.1 Contact stress

Olver (1995) has reported that the micropitting rate increases and the incubation period decreases with increasing load. A special feature of micropitting is that it can occur even at moderate loads, below the pitting endurance limit and, that it can cause damage after short running times (Höhn et al., 1996). Therefore, contact stress is not a strong influence, and high loads are not necessary for micropitting to occur. Micropitting occurs in gears and rolling element bearings. An interesting observation is that the contact stress required for gear failure is about half of that required for the failure of rolling element bearings (Graham et al (1981). Differences in surface roughness (i.e., gear surfaces are rougher) might account for this difference but the sliding experienced in gears must play a significant role. Martensite decay depends on the operating load. Bush et al (1961) have found a threshold contact stress below which no transformation occur.

3.7.2.2 Operating temperature

The operating temperature mainly affects the lubrication conditions (i.e., the lubricant viscosity and the friction coefficient). Micropitting resistance generally decreases with increasing gear tooth temperature. An increase in the operating temperature results in a decrease of the lubricant viscosity and the lubricant film thickness and thus, an increase in contact and micropitting occurrence probability. The temperature also influences the behaviour of the additives. In consequence, an increase in temperature, below a threshold, may improve the lubricant performance due to the action of additives (Höhn et al, 1996). An original work has been done by Seireg (2001) by adding the effect of thermal stresses to the mechanical stresses in order to evaluate the contact fatigue life. Higher speeds reduce the transit time in the contact zone and consequently reduce the depth of heat penetration and the thermally affected zone below the surface. This explains the occurrence of shallow micropitting at high speeds and deep pits at low speeds.

3.7.2.3 Operating speed

High rolling velocity, on one hand has a beneficial influence because it supplies lubricant to the contact and increases lubricant film thickness. High sliding velocity, on the other hand is detrimental because it generates heat. An increase of the operating speed improves the formation of the lubricating film but also increases the operating temperature. Therefore, high operating speeds may promote micropitting. The initiation period of micropitting decreases as the sliding speed is reduced. According to Winter & Weiss (1981) micropitting occurs most readily at speeds in the range of 4-10 m/s. As already mentioned, even at low contact stress micropitting may occur because of the effect of sliding. It has been mentioned in section 2.6, that the predominant sliding direction in gears is transverse to the direction of surface lay, therefore perpendicular to the direction of lubricant entrainment. In these conditions there is an opportunity for a loss of oil from the contact which leads to thinning of the film and ultimately to the film collapse (Evans & Snidle, 1996). Due to local pressure variations within the contact, lubricant leaks into the surface valleys allowing the asperity tops to reach through the lubricant film and collide. The higher the operating speed, the smaller are the size and depth of the pits (Seireg, 2001).

3.7.2.4 Slide to roll ratio

Sliding influences the lubrication conditions. Since sliding has such an important role in micropitting, a general conclusion can be drawn that increasing slide-to-roll ratio increases the probability of micropitting (Graham et al., 1981). At low slide-to-roll ratio micropitting is almost eliminated (Webster & Norbart, 1995). One possible explanation for the reduction in micropitting at lower slide-to-roll ratio is that the friction forces developed in the contact are lower at small slide-to-roll ratios (Graham et al., 1981). An alternative explanation might be given by assuming that the rate of wear is proportional to sliding distance, and hence sliding speed (Webster & Norbart, 1995).

3.7.2.5 Wear debris

The experimental observations of mass loss in rolling/sliding contacts confirm the removal of material. The resultant particles are known as *wear debris*. A fatigue crack propagates to form a pit on the surface and releases wear debris into the lubricant. Wear debris analysis could reveal the different modes of failure. The debris particles generated by contact fatigue processes have irregular shapes (Roylance et al., 2000). Brittle particles explode in the inlet region of the EHL contact and the resultant fragments larger than the lubricant film thickness indent the surfaces (Nelias & Ville, 2000). Owing to the relative motion between a particle and a counterface, sliding friction follows and produces heat, which is shared between the particle and the surrounding surfaces. Frictional heating can be severe, even for small and soft particles. Small and soft particles produce flash temperatures higher than 800°C and lead to local melting and denting the surfaces, which may be the first step in scuffing. Hard particles could produce even higher flash temperatures (Nikas et al., 1998). Due to their very fine scale (few microns) wear debris produced by micropitting are difficult to detect using classical filters, but magnetic filters can capture them. Polishing wear produced by micropitting debris has been reported by Errichello (2003).

The limiting values of the factors that influence micropitting, as they have been reported in literature are as follows:

- Hertzian contact stress $\sigma \geq 500\text{MPa}$
- Number of cycles $N \geq 10^5 ... 10^6$
- Rolling speed $v = 4...10 m/s$
- Slide-to-roll ratio $S > 0.0095$
- Lubricant specific film thickness $\lambda < 2$.

3.8 Micropitting prevention

The fact that micropitting is related to the lubricant film collapse suggests that micropitting can be avoided by increasing the lambda ratio by improving the surface finish, increasing the lubricant viscosity and/or increasing the running speed. However, changing the operating conditions is an unlikely measure since gears are designed for specific applications. The choice of lubricant type is very restricted in some applications such as in the aerospace industry. Surface finish improvement can be achieved by super-finishing processes (i.e., the Abral polishing process mentioned by Olver, 1995) but, usually, these are expensive. Compressive residual stress induced by the shot-peening process has been proven to prolong the gear fatigue life (Townsend, 1992) but does not eliminate micropitting (Guagliano et al., 2002). It appears that other solutions should be found. One possible solution is to deposit a hard tribological coating on the surface of the gear tooth. It is believed (Polonsky et al., 1997) that such a coating can protect the near-surface layers from the contact stress spikes produced by surface roughness, thus inhibiting the near-surface crack initiation. PVD and reactive CVD processes appear to be the most advantageous. The uniform and extremely hard coatings obtained with these techniques do not require an excessive heating of the substrate.

Bull et al. (1997) proposed the following possible ways in which vapour deposited coatings can improve micropitting performance: a reduction of the surface plasticity due to the high hardness, a reduction of the asperity tractions due to the low friction coefficient, a reduction of reactivity between EP additives and steel due to chemical inertness.

Molybdenum disulphide (MoS_2) is widely used, for example, in modern aeronautics and aviation industry as a solid lubricant coating and self-lubricating structural material. The level of adhesion limits its tribological properties. For applications in which a high wear resistance and a low coefficient of friction is required, *carbon-based coatings*, like DLC seem to be the most suitable choice. These films combine a high abrasive and adhesive wear resistance (hardness 1500-3000 HV), with low coefficients of friction (μ=0.2 vs. steel) in non-lubricated operations. Carbon coatings can be deposited easily at temperatures below 200°C, they are chemically inert and can be used up to temperatures

of approximately 400°C. Pure, electrically insulating carbon coatings (diamond like carbon, DLC) have the highest wear resistance (hardness circa 3000 HV) closely followed by metal containing carbon coatings (Me-DLC or Me-C:H), which are electrically conductive. The behaviour of tribological coatings under rolling/sliding conditions has been described in the literature (Nakajima et al., 2000); Straffelini et al., 1999). Nakajima et al. (2000) tested the surface durability of a thermally sprayed WC-Cr-Ni coating in lubricated rolling/sliding conditions using a disc-machine. They found that fatigue life increased with increasing coating thickness.

3.9 Summary

In this chapter the failure modes in gears are reviewed, emphasis being placed on surface contact fatigue phenomena. It also contains a review of the literature concerning micropitting and related phenomena (i.e., phase transformations).

In the next chapter the experimental techniques used to investigate micropitting are presented.

~ Chapter 4 ~

EXPERIMENTAL METHOD

4 EXPERIMENTAL METHOD

4.1 General

This chapter describes the methods used to investigate the micropitting phenomenon. Micropitting experiments have been carried out using a twin disc machine, according to a factorial experimental design in order to assess the influence of seven factors. In order to minimise time, the design adopted was a fractional factorial design $2^{(7-4)}$. The analysis of the affected surface, as well as metallographic investigations, have been performed on both, discs and helical gears tested in similar conditions.

4.2 Gear tooth contact simulation

The test methods used to investigate micropitting in gears can be divided into two groups. The first group comprises tests made on gear pairs and the second group comprises laboratory scale tests made on rotating contacting discs. As discussed in section 2.7.2 the use of the discs is based on the fact that the contact between two gear teeth can be described by the contact between two discs, as shown in Figure 2-18. Webster & Norbart (1995), among others, have shown that the experiments made on roller discs faithfully reproduce micropitting as found on gears. In addition, these types of tests are less expensive and the times to failure are considerably shorter because the time between two loading cycles is reduced. The disc machines reported in the literature are designed to work either with two rollers (Widmark & Melander, 1999; Shotter, 1980; Webster & Norbart, 1995) or three rollers (Berthe et al, 1977; Graham et al., 1980; Olver, 1986).

In this work, the experimental study of micropitting was carried out using a two-disc machine designed at the University of Newcastle, which is schematically represented in Figure 4-1. A complete description of the rig is given elsewhere (Holland, 1997 and

Aylott, 1998). The rig has been designed to produce contacts that are as realistic as possible with respect to operating conditions in gears. Figure 4-2 shows the diagram of forces acting in the two-disc machine. A three-phase motor drives a lower shaft (A'B'C') which is connected to the upper shaft (ABC) via a pair of gears and two pairs of bearings. At the other ends of the two shafts (C and C') the disc samples are fitted. Different loads can be applied at the extremity (C") of the hanger (A"B"C") and the load is transmitted via a pivot (B") to the upper shaft and consequently from the upper shaft to the discs.

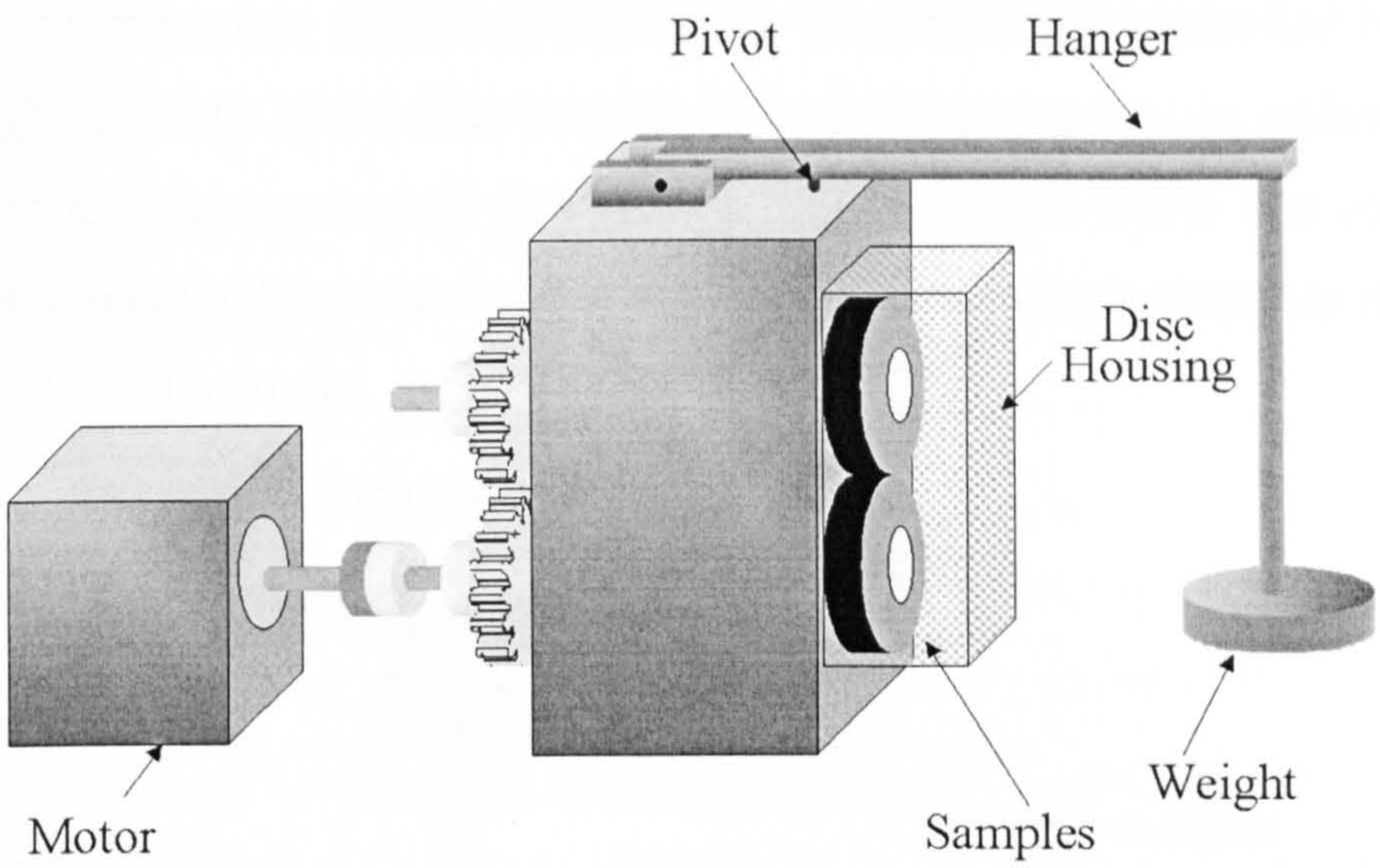


Figure 4-1. Twin disc machine.

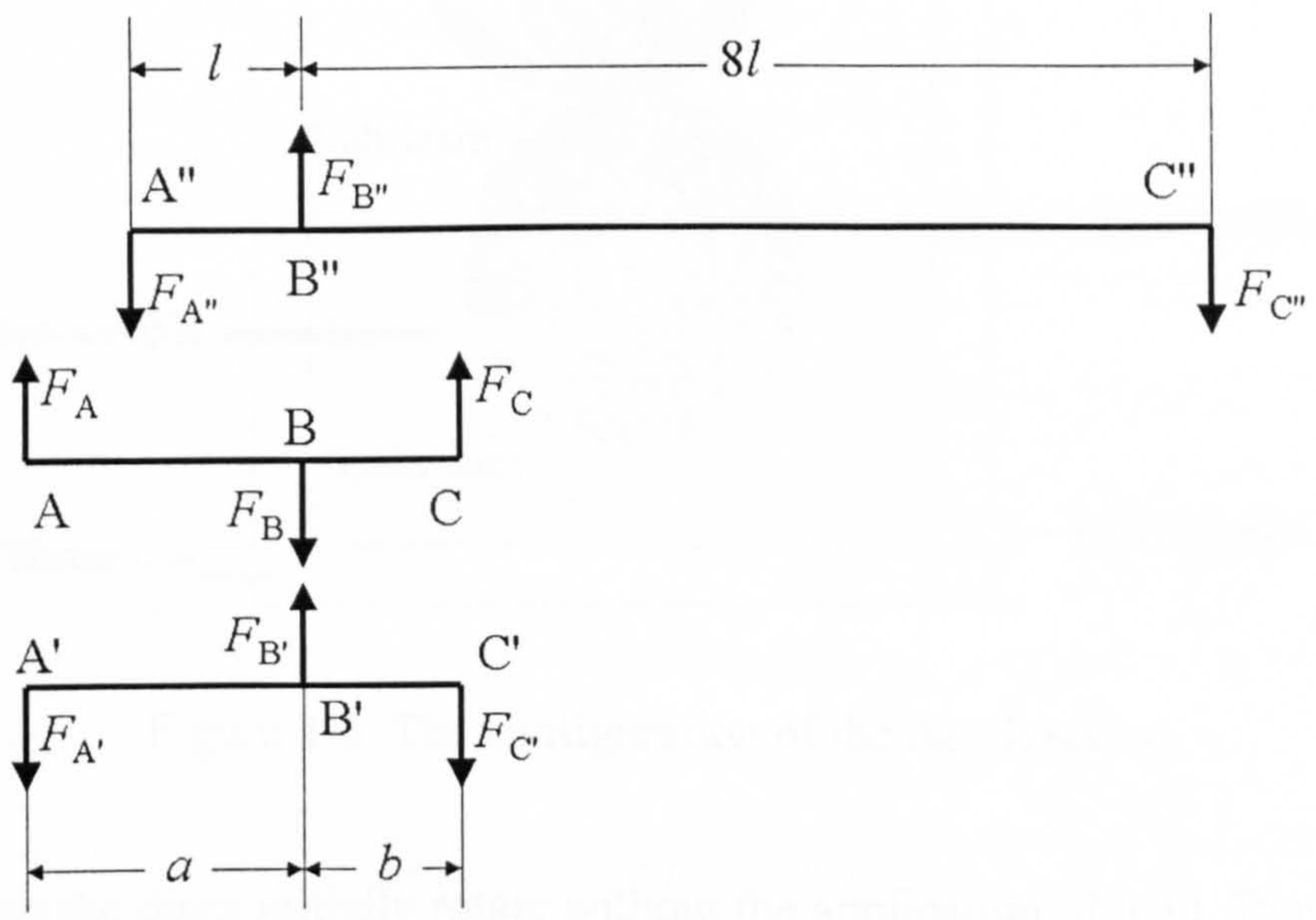


Figure 4-2. Diagram of forces in the twin disc rig. l = 50 mm, a = 90 mm, b = 40 mm.

Disc samples have an inner diameter of 25 mm to fit on the shafts. The sum of the outer diameters is 120 mm. Different slide-to-roll ratios can be achieved by choosing discs of different diameters but keeping constant the sum of the diameters of the two discs.

The discs are enclosed in the disc housing, which contains lubricant and is cooled by water (see Figure 4-3). The lubricant is sprayed in-between the rolling discs by the means of a tube that comes down from the lubricant reservoir. The temperature of the sprayed lubricant is controlled by a cartridge heater situated at the entrance in the disc housing. The temperature of the lubricant in the disc housing is controlled by the water flow through the walls of the disc housing. The disc housing has an orifice where the lubricant can exit the system and flows into a heat exchanger where it is cooled. From the heat exchanger the lubricant flows into a basin from where, it returns to the reservoir due to the action of a pump.

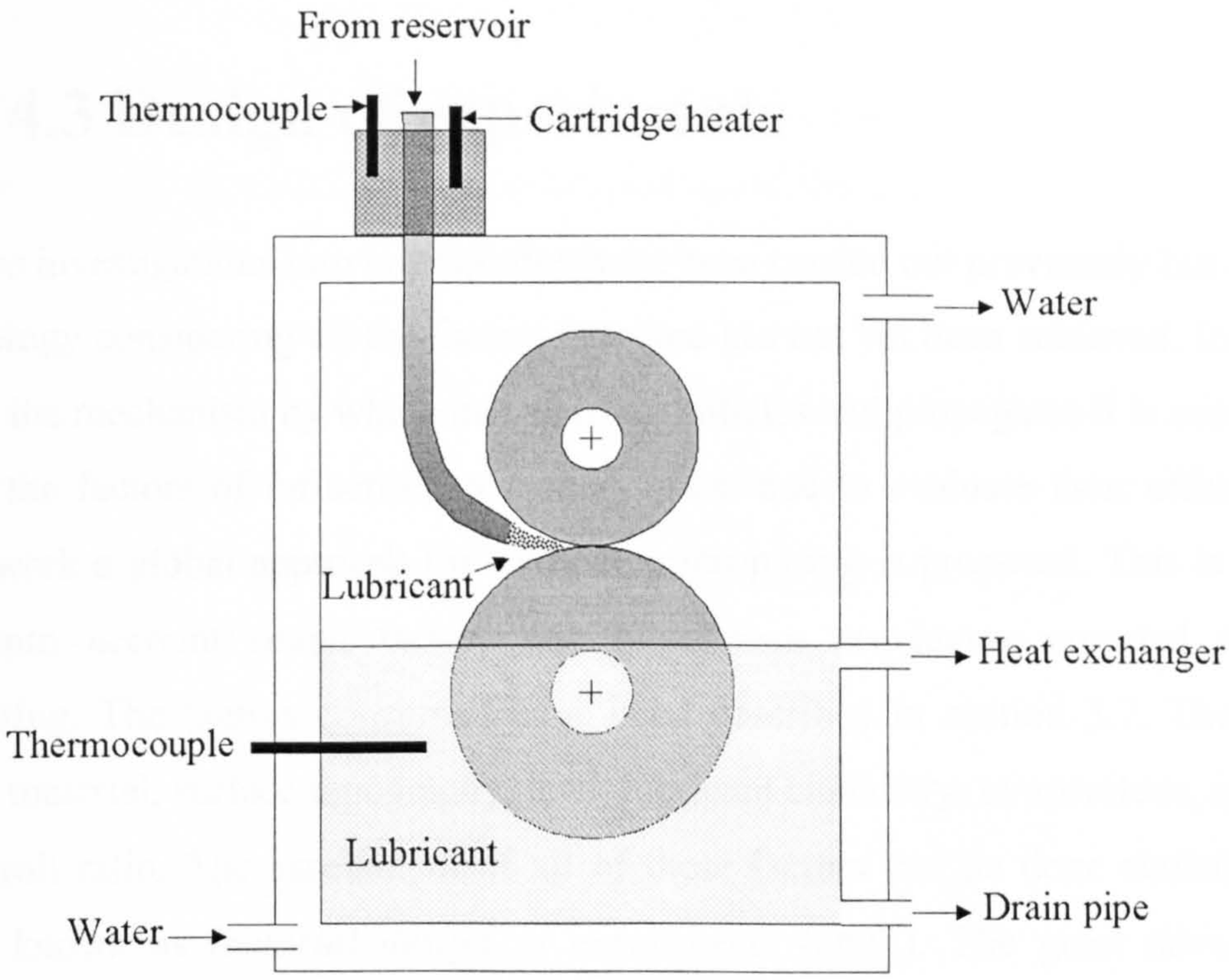


Figure 4-3. The configuration of the disc housing.

During testing the discs initially rotate without the application of load. Once the desired speed has been reached the load can be applied. This is done gradually by adding

weights on the hanger. Different loads can be achieved by loading the hanger with different weights.

Considering Figure 4-2 the load applied can be calculated as

$$P = \frac{9g}{1 + b/a} \cdot m = 61.12 \cdot m \qquad (4\text{-}1)$$

where, $g = 9.81$ m/s^2 is the acceleration due to gravity and $m = 10.16$ kg is the mass of a single weight.

It must be noted that, with this rig, the micropitted specimens cannot be run for a large number of cycles. This is due to the high level of vibrations which lead to the failure of the ball bearings (A and A') and/or needle bearings (B and B'). This means that the rig is more suitable for studying the initiation of micropitting and its propagation in the early stages.

4.3 Design of experiments

Extensive investigations into micropitting have been carried out previously but a unified methodology considering all the factors involved has not yet been achieved. In order to describe the mechanism by which micropitting initiates and propagates it is necessary to identify the factors of influence, to monitor them and to evaluate their effect. In the current work a global approach for studying micropitting is proposed. This is done by taking into account seven factors that have been previously reported to affect micropitting. The factors concerned have been described in section 3.7. They are as follows: material, surface topography, load, lubricant chemistry, temperature, speed and slide-to-roll ratio. The assessment of all of these factors can be done statistically by what is known as *factorial design of experiments* (DOE). The great advantage of factorial experiments is that they allow the experimenter to study a number of factors simultaneously and see how the factors interact with one another. One difficulty with factorial designs is that the number of combinations increases exponentially with the number of factors. In our study, a full factorial design would require $2^7 = 128$ experiments! The *fractional factorial design* enables the study of the main (linear) effects of each factor with a small number of experiments. It does this by using a

fraction of the experiments required for a *full factorial design*. For seven factors, each with two levels, the theoretical minimum number of experiments is eight ($2^{(7-4)}=8$) so the fraction is 8/128. A *main effect* is the effect of one factor averaged across the levels of other factors. An *interaction* is the variation of the differences between means for different levels of one factor over different levels of the other factors. In such a design, some of the factor combinations are excluded and therefore some of the effects will be confounded. Confounded effects cannot be estimated separately and they are said to be aliased. A measure of the aliasing relationships that exist in a design is given by the resolution. A $2^{(7-4)}$ design with eight experiments has a resolution III which means that no main effect is aliased with another main effect and at least one main effect is aliased with at least one two-factor interaction and two-factor interactions are aliased with each other. Complete descriptions of the mathematical apparatus involved in experimental design of experiments can be found in a number of relevant textbooks. The reader is advised to consult Cochran & Cox (1993). The factors involved in this study and their corresponding levels are given in Table 4-1.

Table 4-1. Factors influencing micropitting and their corresponding levels.

Factor		Low level	High level
1	Steel grade	EN 36	SAE 8620H
2	Surface lay	Longitudinal	Transverse
3	Load *P* (N)	2484.07	4968.14
4	Lubricant type	Base oil + 4 % EP	OEP-80
5	Temperature, T(°C)	60	100
6	Speed, *v* (rpm)	1000	1200
7	Slide-to-roll ratio, *S*	0.23	0.33

The computing for experimental design has been done with Minitab 13 software. The resulting matrix of experimental design is shown in Table 4-2.

Table 4-2. The design of micropitting experiments.

Experiment	Steel	Surface lay	Load (N)	Lubricant	Temperature (°C)	Speed (rpm)	Slide/roll ratio
1	EN 36	Longitudinal	2484.07	Mineral oil	60	1200	0.33
2	SAE 8620H	Transverse	4968.14	Mineral oil	60	1200	0.23
3	EN 36	Transverse	2484.07	OEP-80	60	1000	0.23
4	SAE 8620H	Transverse	2484.07	Mineral oil	100	1000	0.33
5	SAE 8620H	Longitudinal	2484.07	OEP-80	100	1200	0.23
6	EN 36	Transverse	4968.14	OEP-80	100	1200	0.33
7	EN 36	Longitudinal	4968.14	Mineral oil	100	1000	0.23
8	SAE 8620H	Longitudinal	4968.14	OEP-80	60	1000	0.33

4.4 Disc samples

The discs were manufactured from two gear steel grades: En36 and SAE 8620H, which are widely used for gear manufacturing. The chemical compositions of the two steels are given in Table 4-3.

Table 4-3. Chemical composition of the steels.

wt. %	C	Mn	Cr	Ni	Si	Mo	S	P
EN 36	0.12	0.55	1.2	3.25	0.28	0.12	0.025	0.025
SAE 8620H	0.20	0.80	0.50	0.55	0.25	0.20	0.04	0.035

Samples were cut from forged bars to a diameter 0.5 mm larger than the nominal size considering that 0.5 mm was going to be removed during the grinding process. The

discs were heat treated with a production load of gears at the Bodycote Heat Treatment Plant, Hurdsfield, Cheshire, UK. After carburising at 925°C for 8 hours the samples were quenched in agitated oil to form a hardened case. In order to increase ductility after quenching the samples were tempered at 150°C for 2 hours. The case depth achieved was 1.5 mm. The mechanical and thermal properties of the hardened case of the two steels, relevant in this study, are given in Table 4-4.

Table 4-4. Mechanical and thermal properties of the steels (from MatWeb.com)

	EN 36	SAE 8620H
Yield strength, σ_y (MPa)	931	1033
Young's modulus, E (GPa)	200	205
Shear modulus, G (GPa)	80	80
Poisson's ratio, ν	0.25	0.281
Heat capacity, k (J/g ·°C)	0.472	0.475
Thermal conductivity, α (W/mK)	51.9	46.6

4.4.1 Grinding

Specimen grinding was carried out on a Cincinnati Tool and Cutter Grinder model No 2 fitted with C-BN grinding discs of a grade used for manufacturing gears. Following heat treatment the discs were first ground on both flat faces and then on the test surface. For studying the influence of surface topography on micropitting four pairs of discs were cylindrically ground and four axially ground. Cylindrical grinding produces longitudinal lay meanwhile axial grinding produces transverse lay. Transverse lay is a better approach to gear contact conditions, considering that, on the gear tooth, the direction of grinding marks and the direction of motion are perpendicular. The surface finish is about R_a = 0.4 μm which is a common value for gear surface finish. A drawing of the arrangement used in grinding is shown in Figure 4-4. For the axial grinding the abrasive wheel was in the form of an internal cone.

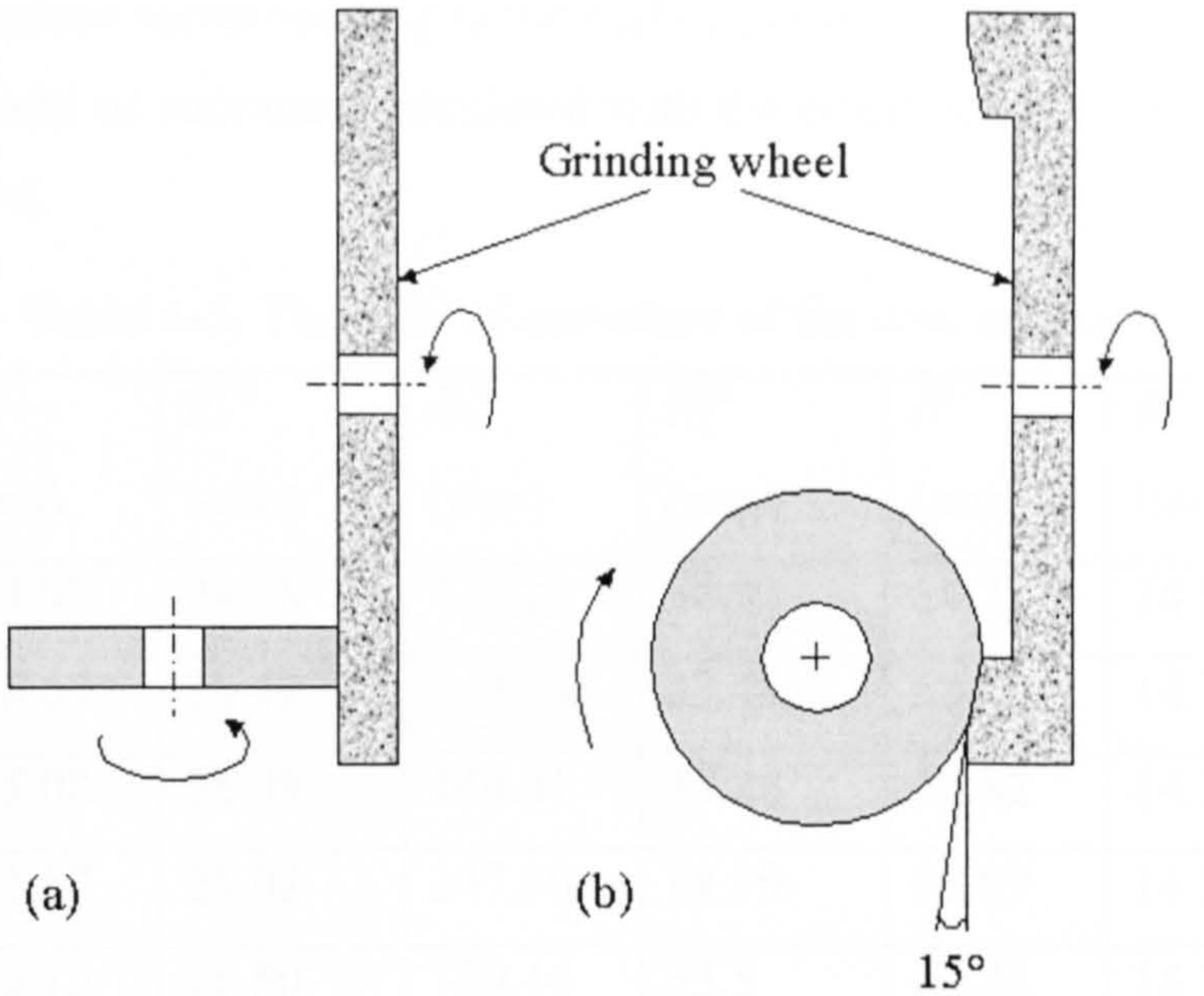


Figure 4-4. The configuration used for grinding. (a) Cylindrical grinding and (b) axial grinding.

The discs, especially those axially ground are crowned. The minimum radius of curvature R'' has been measured using a digital Vernier calliper and the maximum radius of curvature R' has been calculated from the equation of an ellipse, as shown in Figure 4-5.

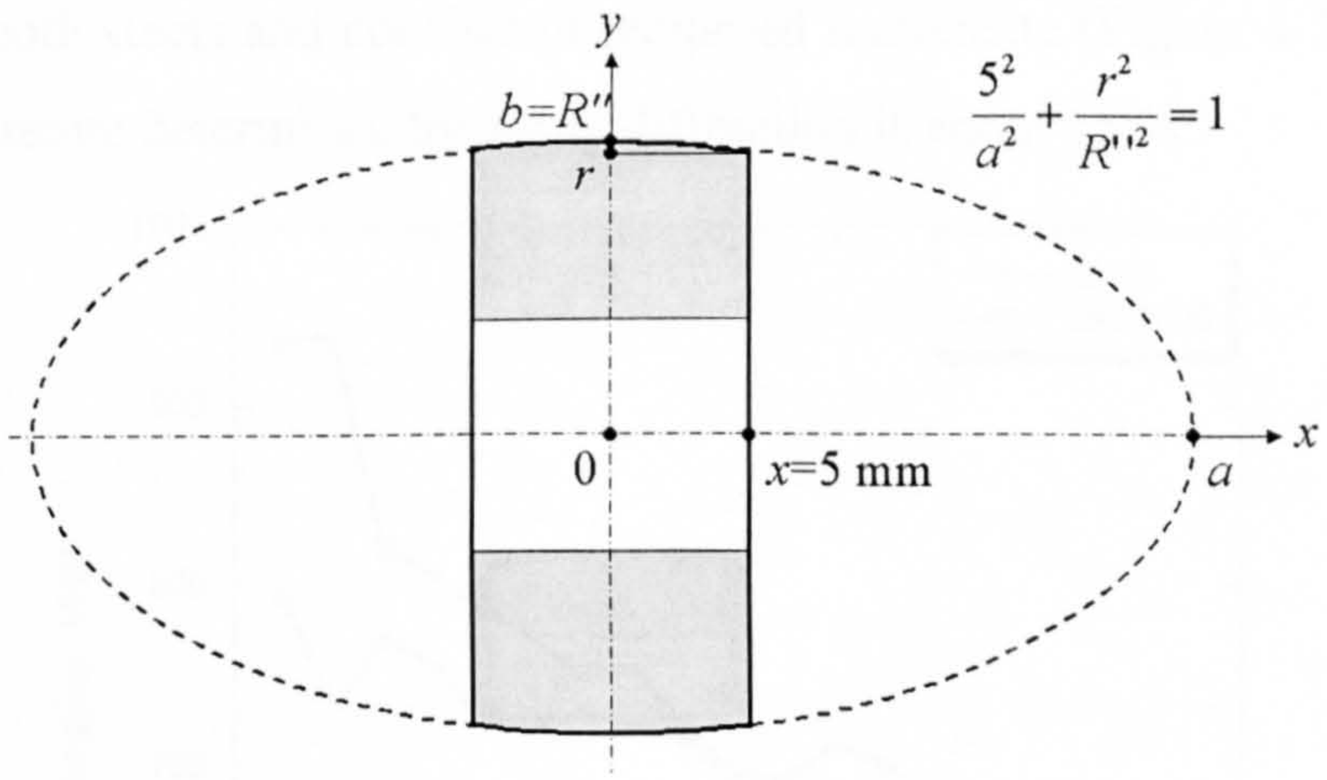


Figure 4-5. Radii of curvature of the disc sample.

From Figure 4-5 the maximum and minimum radii of curvature R' and R'' are given by the following expressions

$$R' = a = \frac{5}{\sqrt{1-(r/R)^2}} \tag{4-2}$$

$$R'' = b \tag{4-3}$$

The radii of curvature corresponding to the eight pairs of test discs and the minor, major and equivalent radii of curvature calculated with the equations 2-23 & 24 and 2-26 are given in Table 4-5.

Table 4-5. The radii of curvature of the disc samples.

Experiment	R_1' (mm)	R_1'' (mm)	R_2' (mm)	R_2'' (mm)	R' (mm)	R'' (mm)	R_e (mm)
1	124.97	24.98	132.25	34.97	64.25	14.57	30.59
2	128.67	26.48	144.70	33.49	68.10	14.78	31.73
3	105.08	26.49	144.41	33.36	60.82	14.76	29.96
4	111.87	25.02	147.90	34.99	63.69	14.58	30.48
5	115.13	26.50	129.44	33.5	60.93	14.79	30.02
6	144.27	24.97	111.78	34.97	62.98	14.56	30.29
7	148.55	26.47	118.13	33.48	65.80	14.78	31.19
8	144.35	25.00	132.33	35.01	69.04	14.58	31.73

The depth of case hardening after grinding is about 1 mm. The hardness profiles corresponding to the two steels are shown in Figure 4-6. The microstructure of the case is similar for both steels and consists of tempered martensite (Figure 4-7). The amount of retained austenite determined by X-ray diffraction is about 15 %.

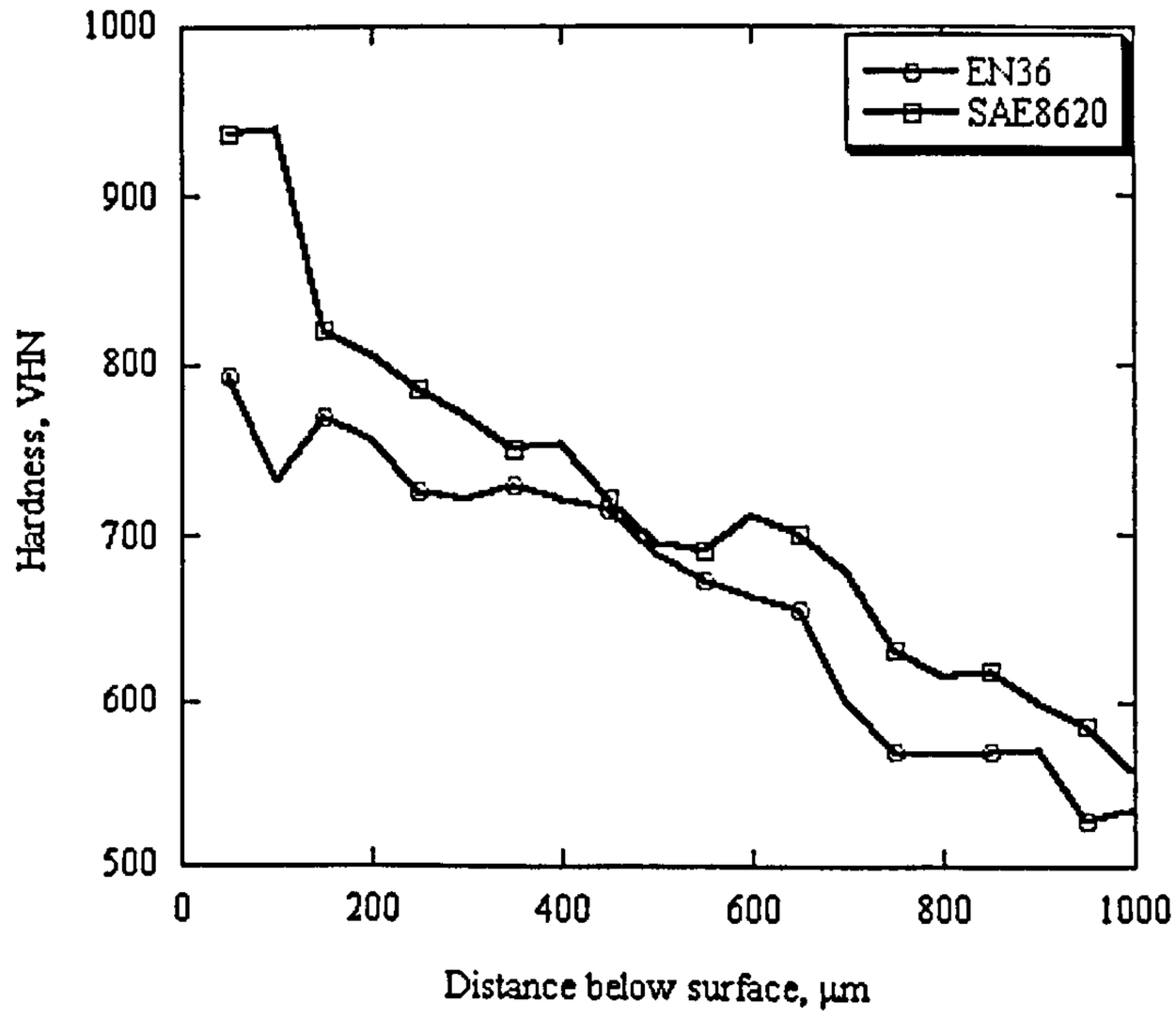


Figure 4-6. Hardness profile after grinding. 300gf load.

Figure 4-7. Microstructure of the hardened case.

4.5 Lubricants

In order to assess the influence of lubricant chemistry, experiments were carried out with two different lubricants, a mineral oil with 4 % EP additives (Anglamol A99) and OEP-80, used for marine gear lubrication. The dependence of dynamic viscosity on temperature has been determined with a Brookfield digital viscometer model RVTD, which is illustrated in Figure 4-8. The viscometer measures the torque required to rotate a spindle immersed in the lubricant. The spindle is driven by a motor through a calibrated spring and deflection of the spring is measured with a rotary transducer, which provides a torque signal which is indicated by a digital display. For a given viscosity, the resistance to flow is proportional to the spindle speed of rotation and is related to the spindle size and geometry. The spindle factors are used to convert the instrument torque, expressed as the display reading, into centipoise. The usual vessel used for viscosity measurements is a 600 ml low form Griffin beaker.

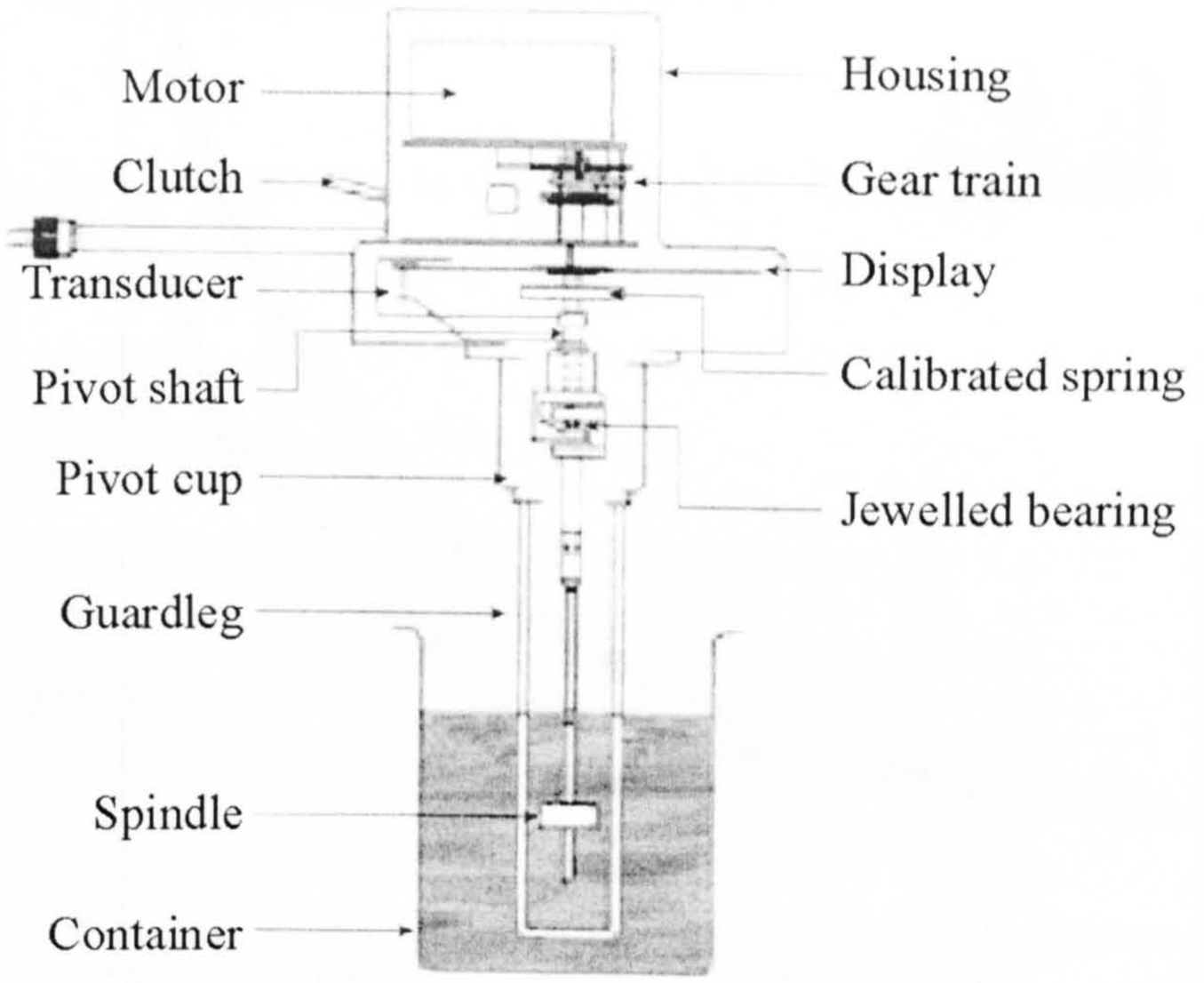


Figure 4-8. Digital Brookfield viscometer model RVTD.

The viscosity has been determined using a spindle RV2 at a rotational speed of 100 rpm which gives a spindle factor $f = 4.0$. The recorded values have been used to plot the curves in Figure 4-9. The best fitting of the data is achieved by a power law. The regression coefficient for both curves is R = 0.99. The regression equations are

for mineral oil

$$\eta = 4.05 \cdot T^{-1.06} \tag{4-4}$$

and for OEP-80

$$\eta = 12.82 \cdot T^{-1.29} \tag{4-5}$$

The dependence of viscosity on pressure is shown in Figure 4-10. The curves have been plotted according to the Barus relationship (equation 2-80). The pressure-viscosity coefficient has been calculated with the Wooster equation (2-82).

Table 4-6. The Barus equation for the two lubricants under testing conditions.

Lubricant	Mineral oil + 4 % EP		OEP-80	
T (°C)	60	100	60	100
Barus equation	$\eta = 0.047 \cdot e^{22.13p}$	$\eta = 0.033 \cdot e^{20.65p}$	$\eta = 0.065 \cdot e^{23.49p}$	$\eta = 0.030 \cdot e^{20.36p}$

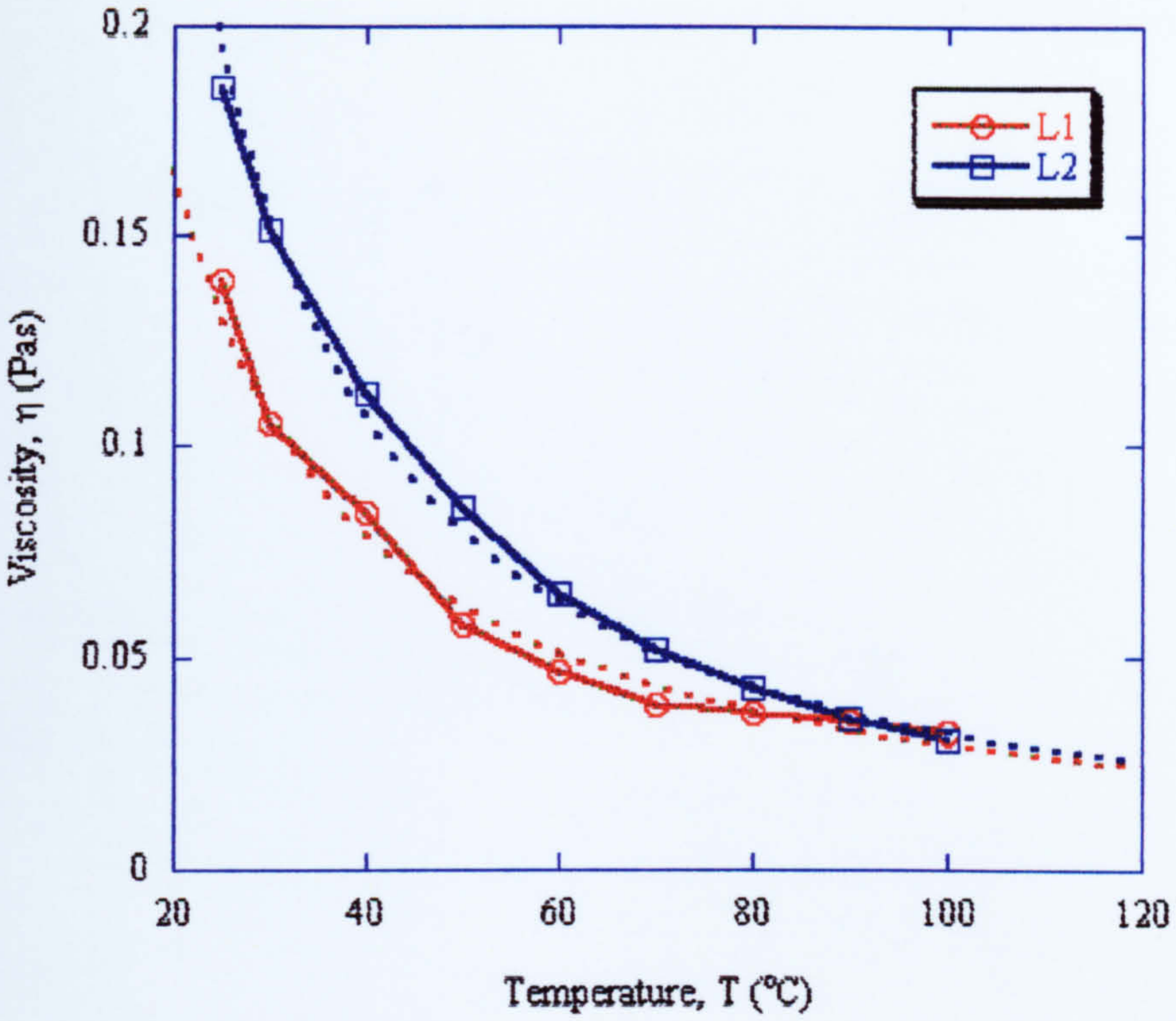


Figure 4-9. The variation of dynamic viscosity with temperature. Dotted lines represent curves fit by a power law. L1 = mineral oil, L2 = OEP-80.

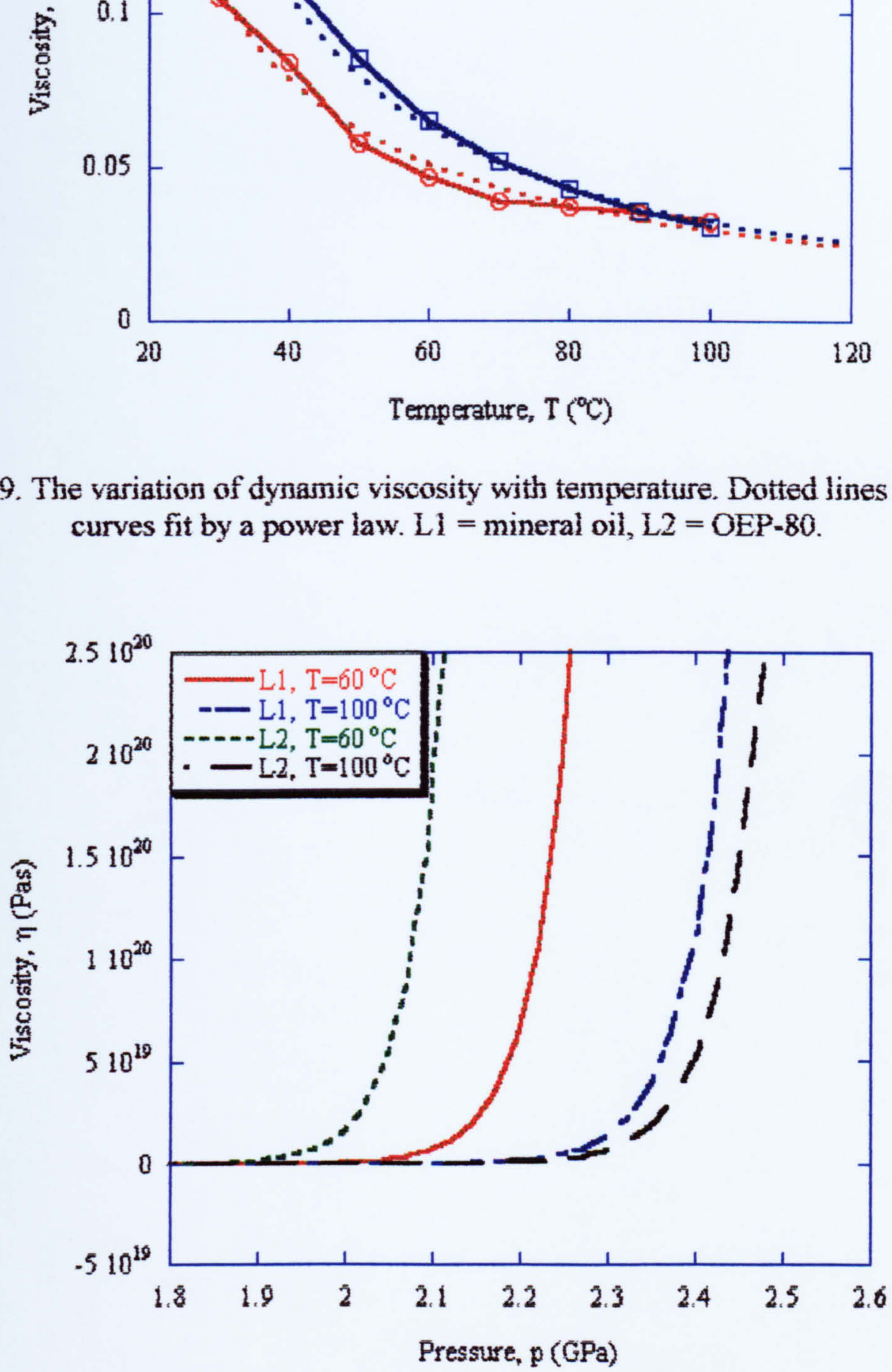


Figure 4-10. The variation of viscosity with pressure (Barus dependence). L1 = mineral oil, L2 = OEP-80.

4.6 Operating conditions

The parameters that describe the disc contact conditions in the eight experiments are shown in Table 4-7. They have been determined using Hertzian theory for elliptical point contact, outlined in section 2.7.3.2 and the Hamrock-Dowson equations (2-92 & 93) for lubricant film thickness in elliptical contacts.

Table 4-7. Operating conditions in the eight micropitting experiments.

	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8
p_0 (GPa)	1.800	2.277	1.826	1.857	1.876	2.284	2.240	2.278
a (μm)	1330	1697	1291	1306	1274	1660	1692	1713
b (μm)	494	613	502	488	496	625	625	607
δ (μm)	21	33	21	21	21	33	33	33
z (μm)	329	407	334	325	329	415	415	404
τ_{max} (MPa)	585	740	593	603	609	742	728	740
u (m/s)	3.76	3.76	3.13	3.13	3.76	3.76	3.13	3.13
G	2361	2463	2506	2298	2266	2172	2203	2614
U	$5.42 \cdot 10^{-11}$	5.01	6.36	3.05	3.47	3.59	3.11	5.78
W	$2.48 \cdot 10^{-5}$	4.43	2.59	2.40	2.47	5.07	4.78	4.43
η (Pas)	$9.3 \cdot 10^{15}$	$6.5 \cdot 10^{19}$	$1.5 \cdot 10^{17}$	$4.5 \cdot 10^{14}$	$2.4 \cdot 10^{14}$	$8.5 \cdot 10^{17}$	$1.7 \cdot 10^{18}$	$1.8 \cdot 10^{21}$
h_{cen} (μm)	0.727	0.699	0.826	0.488	0.524	0.500	0.471	0.791
h_{min} (μm)	0.252	0.237	0.295	0.169	0.187	0.174	0.161	0.264

In Table 4-7 p_0 is the maximum contact pressure, a and b the semi-axis of the contact ellipse, δ the approach of centres, z the depth of maximum shear stress, τ_{max} the maximum shear stress, u the entraining velocity, G, U, and W are the elastohydrodynamic non-dimensional parameters, α the pressure-viscosity coefficient, η the lubricant viscosity and, h_{cen} and h_{min} the central and minimum lubricant film thickness.

4.7 Micropitting investigation

Investigations into micropitting were carried out on the running surface and on cross sections of the disc specimens. Several non-destructive and destructive techniques were employed to monitor the surface condition before, during and after testing. The "during testing" monitoring has been performed after each stage of testing. One stage of testing is defined as $1.8 \cdot 10^5$ cycles. Tests were interrupted after $N = n \cdot 1.8 \cdot 10^5$ cycles where n = 1, 2, 3 and 4. Considering the limitations imposed by the size of the samples the non-destructive techniques used to monitor the surface condition during running were light microscopy, optical profilometry and X-ray diffraction. The Vickers micro-indentation technique was also used to detect any possible changes in hardness occurring on the surface. The same techniques were used after the tests were stopped. In addition to the above techniques, after testing, the surface was observed with scanning electron microscopy and nanoindentation testing has been performed. Metallographic investigations were carried out on longitudinal and transverse sections taken from the specimens using light microscopy and scanning electron microscopy. Nanoindentation testing was also performed on the longitudinal sections. In addition, after each testing stage, the weight and diameters of the disc samples were measured. Similar techniques have been used to analyse micropitting in gears tested according to FZG standard (BS 14635-1: 2000).

4.7.1 Sample preparation

After testing, the disc samples were prepared for metallographic and other investigations. Test surfaces have been observed in the light and electron microscope without any other special preparation. The microstructure has been observed in two locations, by sectioning the disc in longitudinal and transverse direction as shown in Figure 4-11.

The detail at the edge of the specimen is of the greatest importance in this study. The depth of focus in light microscopy is significantly reduced if the surface is curved rather than flat (Vander Voort, 2000). Therefore the flatness at the edge must be nearly perfect. One method for edge retention is to plate the surface of interest with a

compatible metal (i.e, nickel). In this way the rounding will occur at the outer surface of the plating during polishing not at the interface between the plating and the specimen. In order to ensure the edge retention during grinding and polishing, the curved surface of the specimen has been nickel plated in a Watts nickel bath with the following composition:

Nickel sulphate $NiSO_4 \cdot 7H_2O$	240 g/l
Nickel chloride $NiCl_2 \cdot 6H_2O$	20 g/l
Boric acid H_3BO_3	20 g/l

The other sides of the specimens were masked with lacquer.

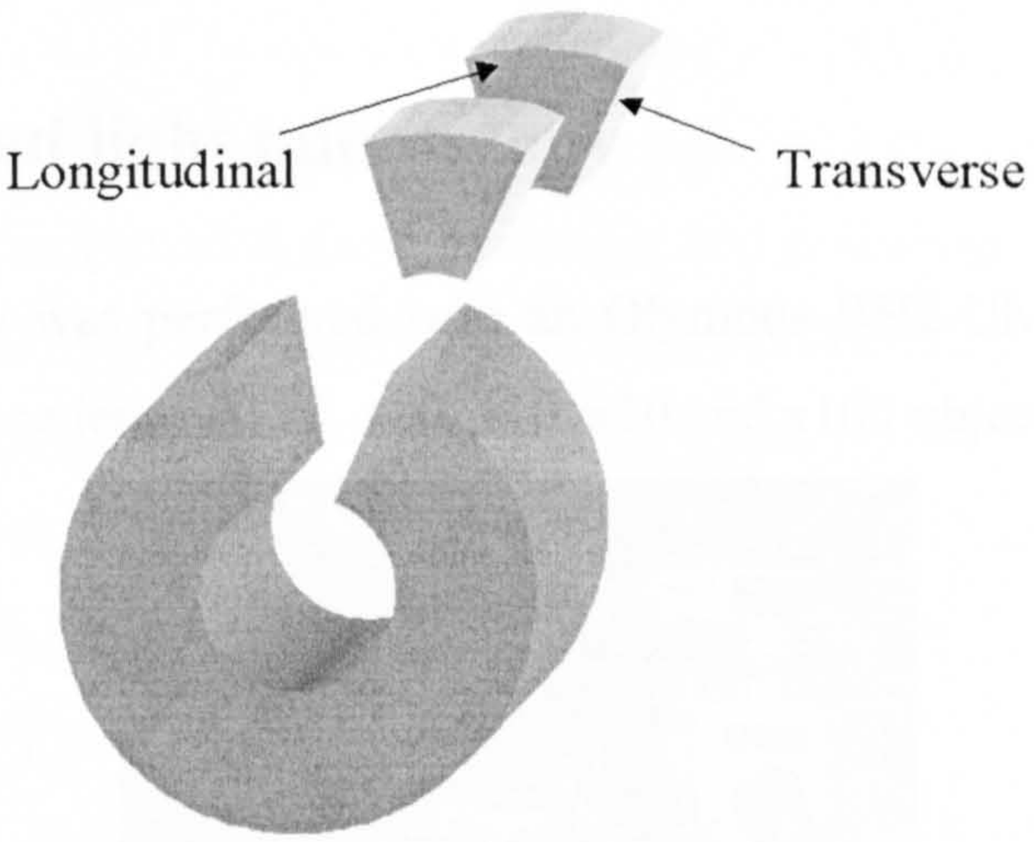


Figure 4-11. Locations of metallographic investigations: longitudinal and transverse.

Specimens were mounted in carbon filled epoxy resin in order to avoid the occurrence of the charging effect during SEM investigation. If the electrical conductivity of the mounting resin is poor the charge injected by the beam accumulates in the specimen and its surface potential rises. This results in high distortions of the image.

Grinding was carried out successively on 240 and 600 SiC papers lubricated by water and then the specimens were polished on alumina cloths of 9, 3 and 1 μm size.
Polished specimens were used to observe cracks and non-metallic inclusions.

In order to reveal the microstructure two types of etchants have been used: *nital* and *picral*. The products of martensite decay have been described in Chapter 3. The dark etching regions (DER) are only revealed by the nital etch (Österlund & Vingsbo, 1980)

whereas the best resolution for the white etching bands (WEB) is achieved by etching with picral solution (Gentile et al., 1965). Both etchants are commonly used in the metallography of steel. Nital is a solution of nitric acid (HNO_3) in methanol.

The 2 % solution is most common, but 5-10 % solutions are used for high-alloy steels or for macro-etching. Picral is a solution of 4 g picric acid in 100 ml ethanol. Nital etches ferrite grain boundaries and ferrite-cementite grain boundaries, while picral attacks the phase boundary between ferrite and cementite or other carbides but not the boundaries between ferrite grains (Vander Voort, 1984).

Electrolitic polishing has been used to remove successive layers from the surface for residual stress and retained austenite profiling by X-ray diffraction technique.

4.7.2 Reflected light microscopy

The light microscopy was performed with an Olympus BH2-UMA (see Figure 4-12) fitted with x10 eyepiece lens and x5, x10, x20, x50 and x100 objective lenses.

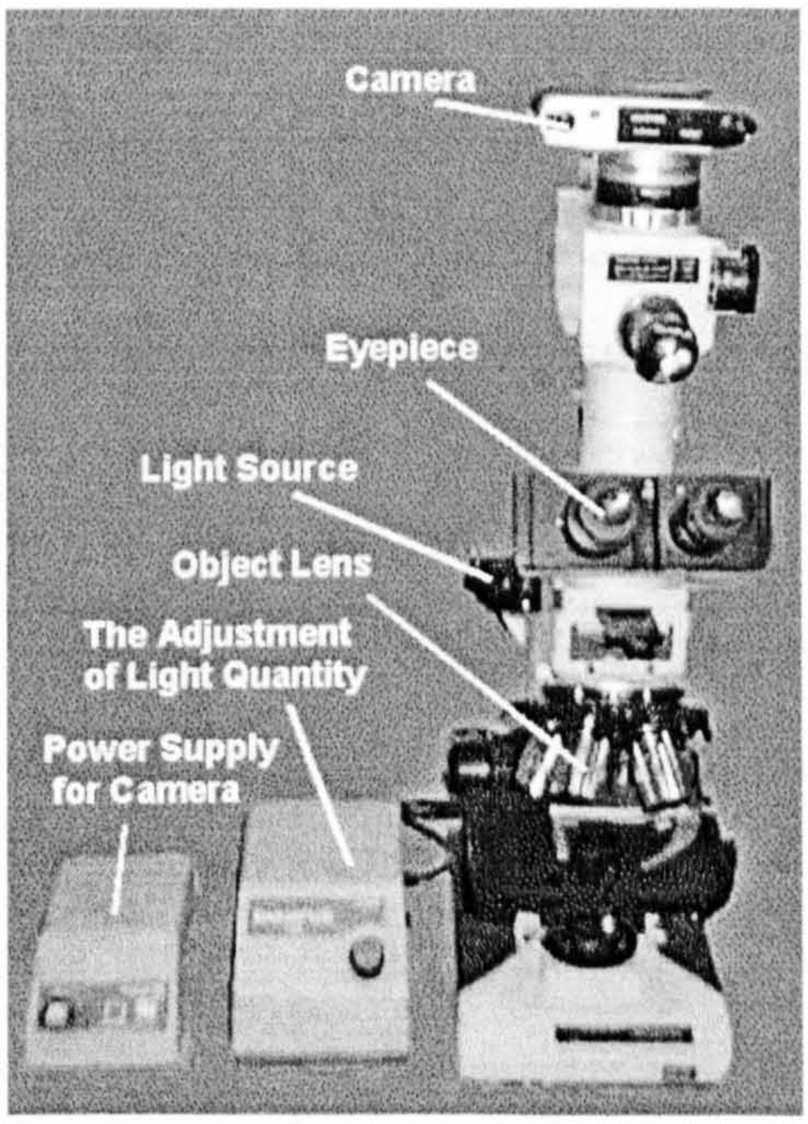


Figure 4-12. Olympus BH2-UMA microscope.

Images can be photographed by a camera attached to the microscope. A video camera is also attached to the microscope and the images are visualised on a JVC video monitor model TM-210PS and output to a personal computer using a video capture card. The obtainable magnifications are listed in Table 4-8.

Table 4-8. Magnification range obtainable with Olympus BH2-UMA microscope.

	Through eyepiece	Camera	Video monitor
Object lens x5	x50	x25	x200
Object lens x10	x100	x50	x400
Object lens x20	x200	x100	x800
Object lens x50	x500	x250	x2000
Object lens x100	x1000	x500	x4000

Light microscopy has been used to observe the test surface after each testing stage and also the cross sections (longitudinal and transverse) prior to and after etching.

4.7.3 Scanning electron microscopy

The scanning electron microscope (SEM) is one of the most versatile instruments available for the microstructural examination and topographical analysis of solids. The SEM used in this study is a CamScan Series 4 manufactured by Cambridge Scanning Co. Ltd. The principle of operation is schematically shown in Figure 4-13. An electron gun (a filament cathode of lanthanum hexaboride (LaB_6) heated to thermionic emission) produces an electron beam of less than 50 μm diameter. The image of the source is demagnified by the condenser lenses and the objective lens focuses the demagnified image of the source onto the sample surface. The electron probe can be less than 5 nm diameter. The probe is deflected by the scan coils over the specimen surface in a rectangular raster. The secondary electrons produced are attracted to a scintillator disc, which is coupled to a photomultiplier. The light emitted by the scintillator passes to the photo cathode of the photomultiplier. After processing this video signal modulates the brightness of the display cathode ray tube. The spot in the cathode ray tube is scanned in synchronisation with the scanning of the electron probe across the specimen. The image resolution can be as high as 5-6 nm and the depth of focus is at least 300 times greater than a light microscope. A large range of magnification can be achieved from less than 10x to greater than 300,000x. The correct environment for operation of the electron optical system is achieved by a high speed vacuum system to a pressure less than 10^{-5} torr.

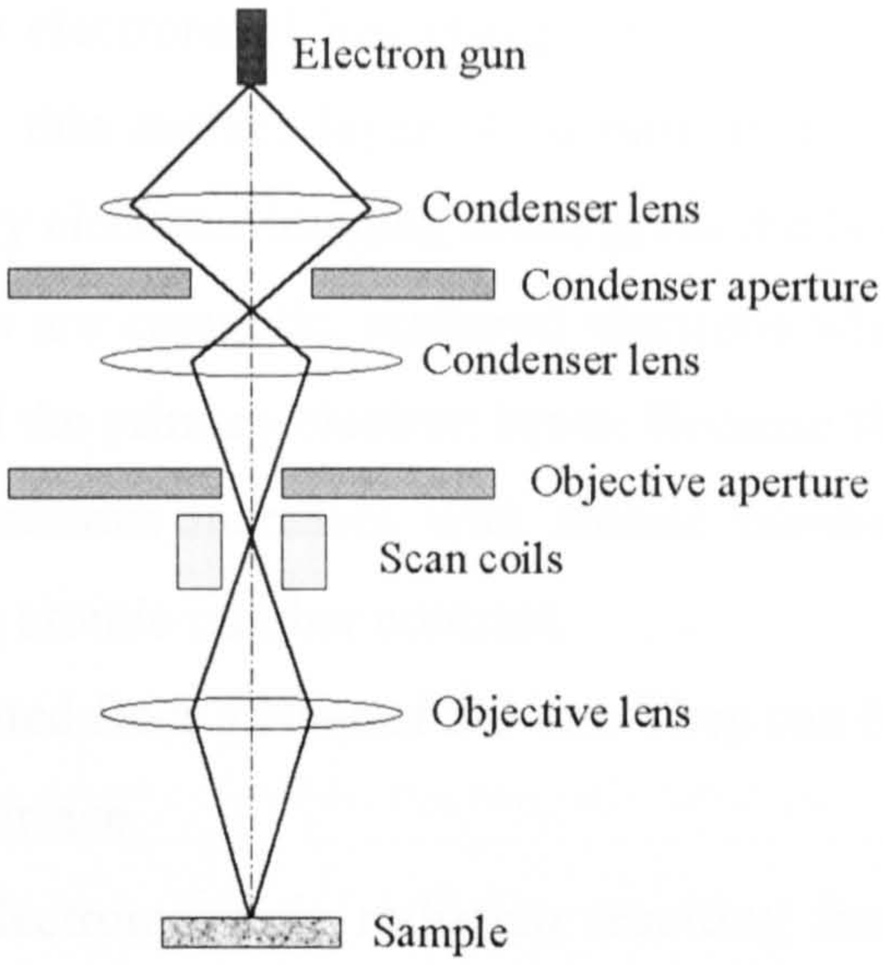


Figure 4-13. Schematic representation of the SEM optics.

The incident electron beam can interact with the specimen in a number of different ways (see Figure 4-14). Elastic scattering affects the trajectories of the beam without significantly altering the kinetic energy of the electron. Backscattered electrons resulting from the elastic events represent an important imaging signal in SEM. Inelastic scattering results in a transfer of energy from the beam electrons to the atoms of the specimen, leading to the generation of secondary electrons, Auger electrons, characteristic and continuum x rays, cathodoluminiscence, etc.

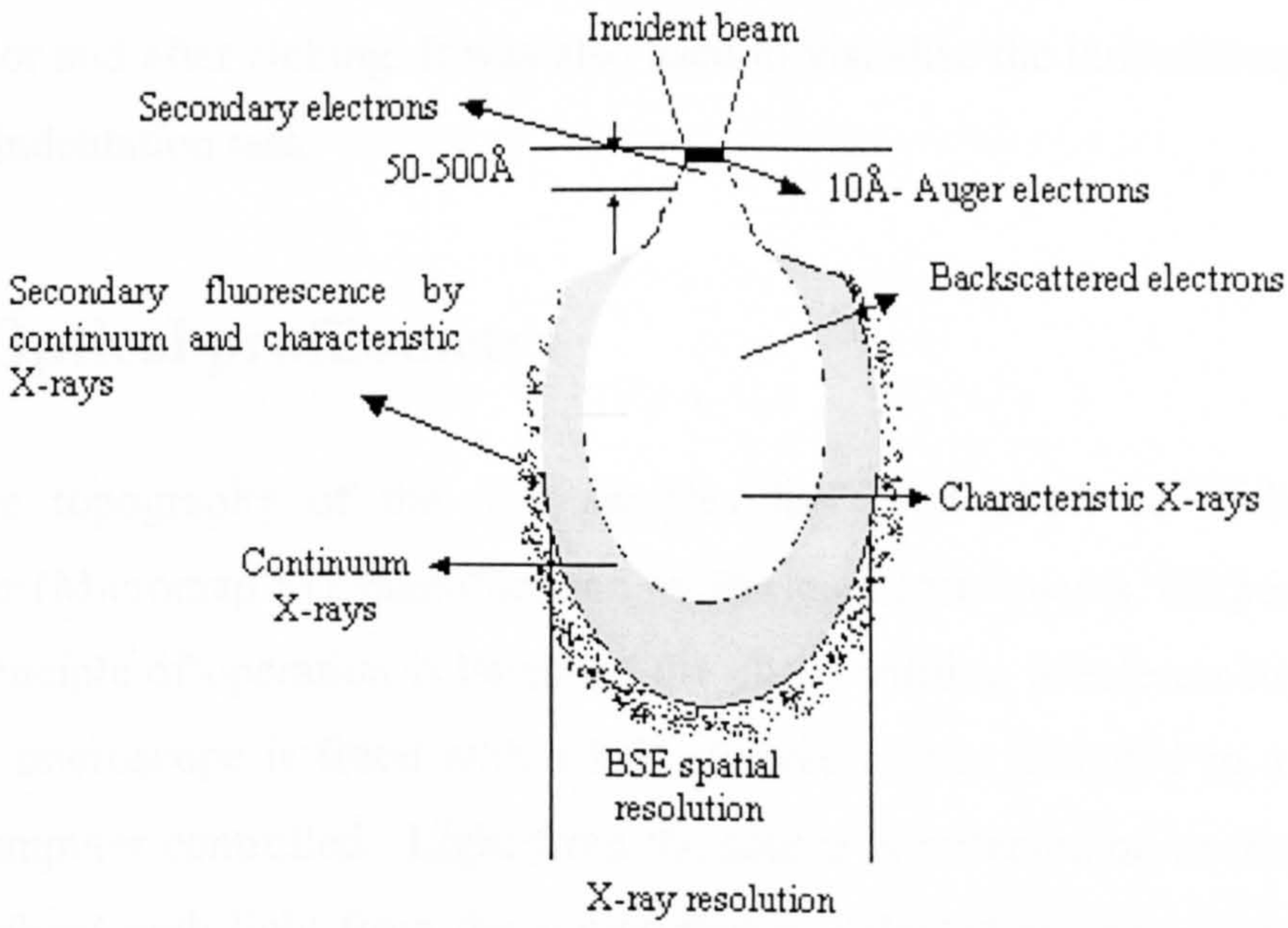


Figure 4-14. Electron interaction with the specimen.

Secondary electrons are electrons of low energy (less than 50 eV) which are emitted predominantly from the thin surface layer (1-10 nm) of the volume irradiated by the primary beam. Secondary electrons imaging mode gives the best resolution.

Back scattered electrons are energetic, scattered electrons which can leave the sample from a large area around the primary electron beam. Because the back scattering ratio of reflected to incident electrons increases with atomic number this imaging mode is widely used for showing atomic number contrast.

Auger electrons are emitted from a layer of 2-3 nm. They can be used for the analysis of the atomic layer at the surface.

X *rays* represent the electromagnetic radiation resulting from inner orbital electron transitions or deceleration of high-energy electrons. Energy Dispersive X-ray Microanalysis (EDX) is a microanalytical technique that uses the characteristic spectrum of X-rays emitted by the specimen after excitation by high-energy electrons to obtain information about its elemental composition.

Cathodoluminiscence is the light emission associated with the excitation of materials by an electron beam. The optics of a scanning electron microscope can be utilized to produce a focused beam and to excite a small region of the sample.

The operating modes used in this study are secondary and backscattered emission. SEM has been used after the tests were stopped to observe the test surface as well as on cross sections prior and after etching. It was also used to visualise the indentations performed in the nanoindentation test.

4.7.4 Optical profilometry

The surface topography of the disc samples has been observed with an optical profilometer (Micromap 512 manufactured by Burleigh Instruments, Harpenden, Herts., UK). Its principle of operation is based on the phase shifting interferometry (Bhushan, 2001). The microscope is fitted with a half silvered mirror mounted in a piezo drive which is computer controlled. Light from the source is reflected on to the test surface and is interfered with light from the source that is reflected by the mirror. When two coincident light waves interfere with each other, they form a characteristic light and

dark fringes. The spacing of the fringes corresponds to areas of different heights on the sample's surface. The light source is filtered to provide illumination of a precisely known wavelength. The information contained within the interference pattern is deconvoluted by the phase-shifting technique. A complete description of the optical profilometer is given in McGurk (1997). An interference microscope compares the sample surface to an internal reference mirror. To measure the sample's relative height from the interference pattern, a piezoelectric crystal moves the reference mirror in small increments (90 degree segments).

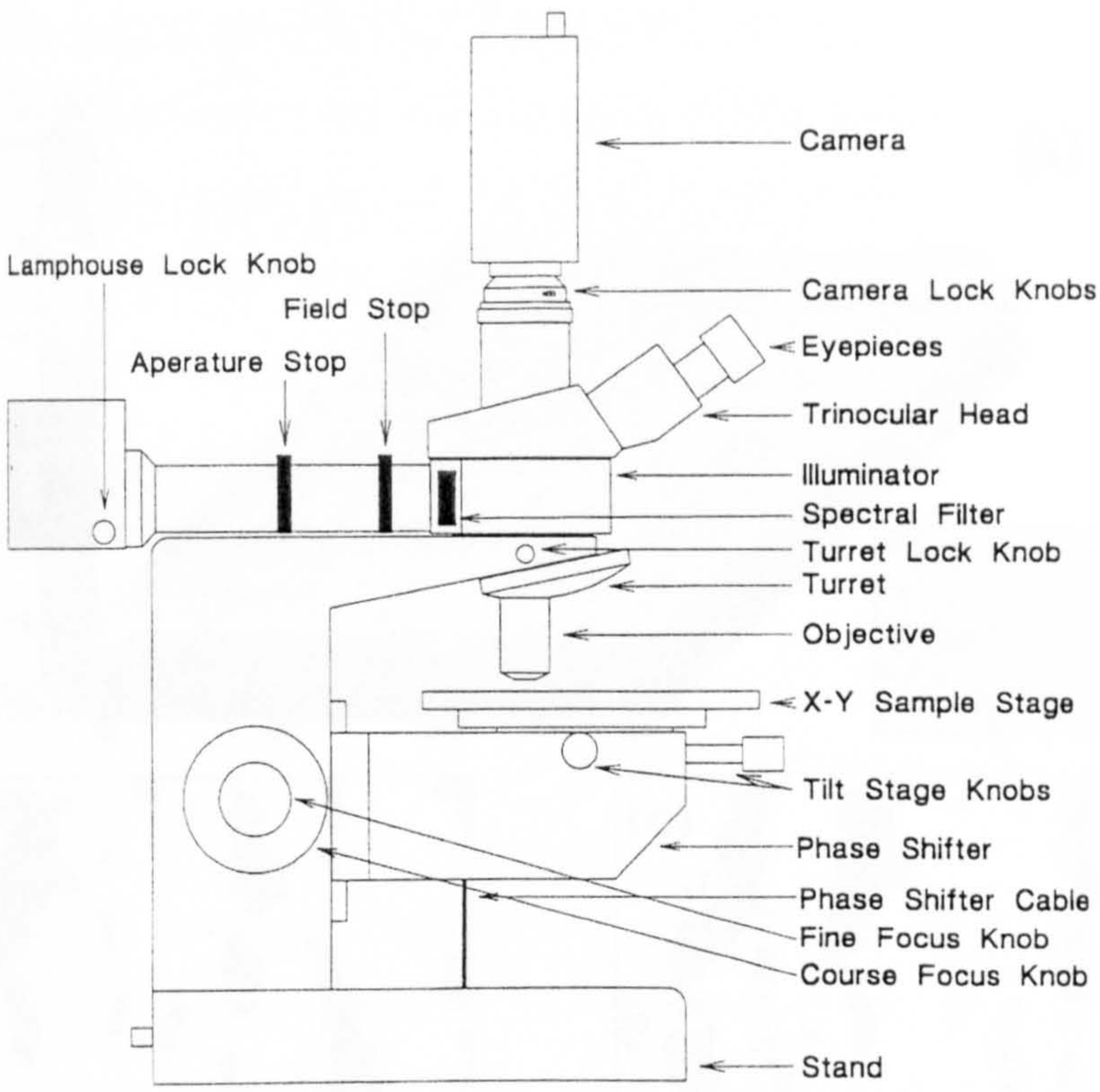


Figure 4-15. Optical profilometer.

Table 4-9. Objectives specifications.

Magnification	x2.5	x10	x40
Measurement area (μm)	3000x2500	750x640	190x160
Sampling interval (μm)	5.4	1.35	0.34
Resolution (μm)	3.7	1.1	0.55
Working distance (mm)	11.1	4.0	3.8
Depth of focus (μm)	50	4.4	1.1
Interference depth (μm)	40	20	4

The computer stores an image of the sample at each position then calculates the phase equation, which determines the exact location of each pixel in the camera image in terms of its height above the surface. The specifications of the objectives are given in Table 4-9. The Micromap 512 profilometer can operate in two modes referred to as wave and phase mode. The wave mode is suitable for rougher surfaces characterised by a peak-to-valley P-V< 80 μm whereas the phase mode is more accurate and it can be used to characterise surfaces with P-V< 3 μm. In the current work the profilometer has been operated in the wave mode. Wave mode measurement uses a broadband filter of 5610 Å.

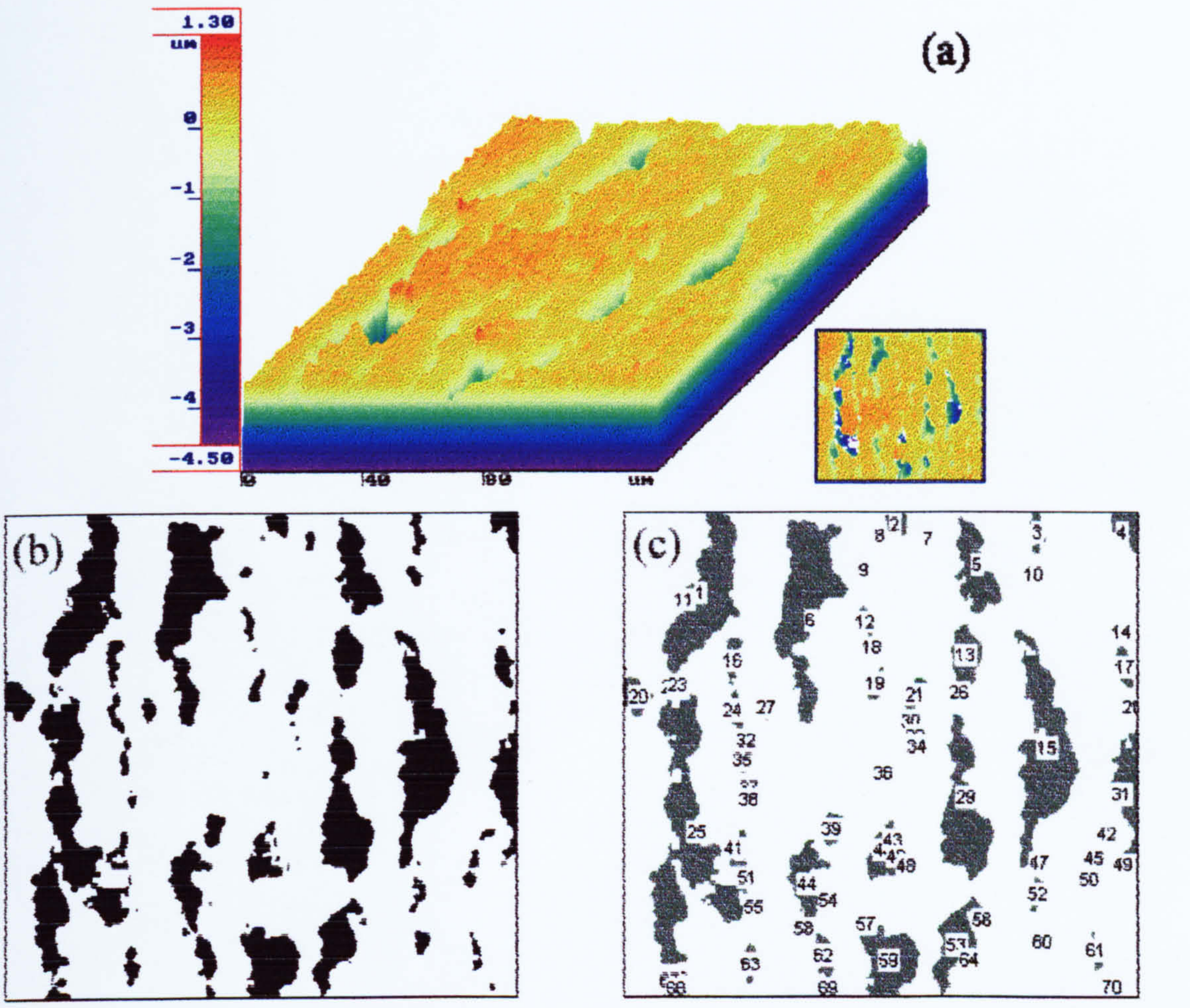


Figure 4-16. Image processing and micropitting measurement. (a) Surface image captured by optical profilometer; (b) Surface contour image after processing; (c) Analysed image; micropits are counted, measured and labelled using Scion Image. The results are as follows: number of micropits = 70, mean area = 53.9 μm^2, minimum area = 0.19μm^2, maximum area = 616 μm^2, total micropitting area = 20.8 %.

The percentage of micropitting (M) on the running surface has been measured by processing the surface contour maps captured with the optical profilometer. The total area of micropitting, the number of pits and the mean area of the pits have been determined using Scion Image software. Figure 4-16 shows a sequence of image processing and micropitting measurements. The Micromap 512 software allows the selection of colours according to the surface height. After a sufficient number of cycles, the grinding marks are completely worn out. By selecting the same colour for all the points below the reference line it is possible to obtain an image of the micropits. More attention must be paid if the grinding marks are still visible. In this case, after colouring the points below the reference line, the micropits and the remaining valleys will appear in the same colour. Since pits and valleys have very different geometry the valleys can easily be removed from the image.

4.7.5 Microindentation

A Shimadzu microhardness tester (type-M, Figure 4-17) was used to monitor the hardness of the test surface. The specifications of the Shimadzu tester are listed in the Table 4-10. The indentation of the disc surface was performed with 25 gf load applied for 15 seconds. This low load has been chosen in order to avoid the presence of big indentation marks on the surface since they can influence the micropitting phenomenon. It is very difficult to accurately measure the diagonals of such small indents using the Shimadzu microscope. For this reason the indents have been measured with the optical profilometer.

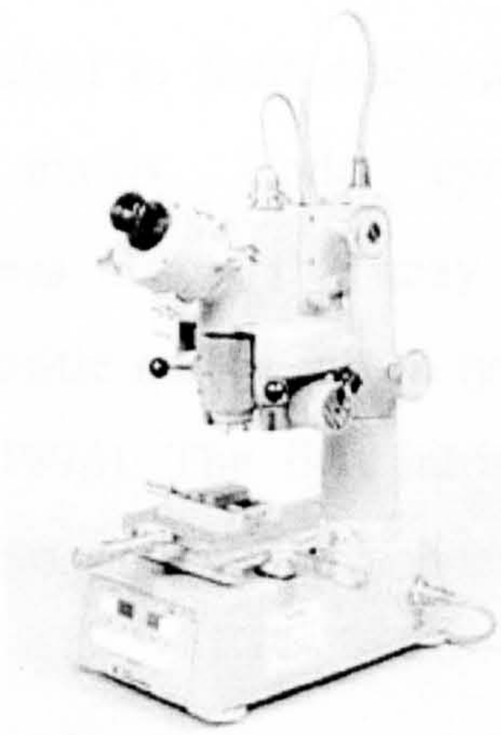


Figure 4-17. Shimadzu microhardness tester.

Table 4-10. Shimadzu microhardness tester. Specifications.

Periods of loading (s)		5, 10, 15, 30, and 45
Loading capacities (gf)		15, 25, 50, 100, 200, 300, 500, and 1000
Indenter		Vickers diamond pyramid
Microscope magnifications		x100 and x400
Micrometer ocular	Maximum scale (μm)	200
	Standard scale (μm)	20 one division
	Measuring scale (μm)	0.5 one division

4.7.6 Nanoindentation

The case microstructure of a carburised, quenched and tempered steel consists of tempered martensite, retained austenite and carbides. These regions are sufficiently small (<10 μm) that only nanoindentation can be used to reliably measure their properties. Moreover, in fatigued specimens phase transformations take place. The mechanisms of these transformations as well as the properties of the new phases are not very well understood, yet. Because of these complex microstructures containing different phases the classical method of measuring Vickers microhardness at moderate load (usually ≈ 3N) gives only an average material response which does not fully characterise the surface. Nanoindentation allows more localised measurements that directly relate to the mechanical properties of individual microstructural features.

The detrimental role of microheterogeneities (non-metallic inclusions, primary carbides) in gear steels is generally ascribed to their elasto-plastic behaviour, which may be different than that of the matrix. Under cyclic loading conditions, these microheterogeneities act as stress raisers that may be responsible for local plastic deformation, decay of the martesitic matrix into a type of ferritic structure and crack nucleation (Lamagnere et al., 1996). The first intrinsic modulus and hardness data measured directly on microheterogeneities have been published by Lamagnere et al (1996).

Table 4-11. Mechanical properties of microheterogeneities (Lamagnere et al., 1996)

		H (GPa)	*E* (GPa)
Al_2O_3		32.2	375
TiN		21.5	380
$(Al_2O_3)_3(MgO)$		26.6	279
$(Al_2O_3)_6(CaO)$		18.2	195
$(Al_2O_3)_2(CaO)$		9.6	126
MnS		3.4	103
MnCaS		3.2	107
Mo_2C		20.4	302
VC		23.9	320
Martensite	AISI 52100	8.3	197
	M50	7.7	214

In the nanoindentation method, the hardness and Young's modulus are calculated from the load-displacement data. A typical indentation load-displacement curve is shown in Figure 4-18.

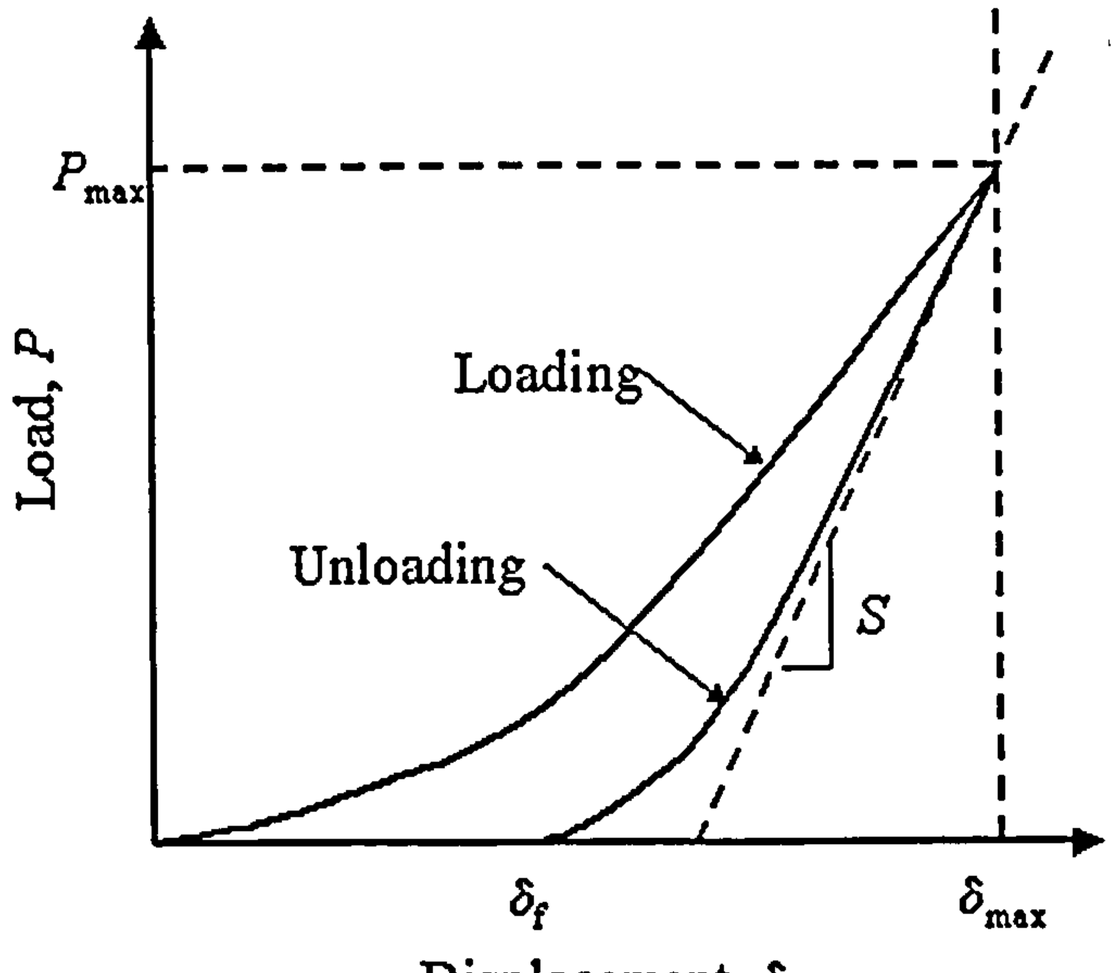


Figure 4-18. Typical load-displacement curve.

The Oliver and Pharr method, which is based upon relationships developed by Sneddon (1965) for the penetration of an elastic half space by indenters that can be described as

solids of revolution of a smooth function gives equations (4-6)–9 for hardness and Young's modulus calculation.

$$H = \frac{P_{\max}}{A_c} \tag{4-6}$$

$$E_r = \frac{\sqrt{\pi}}{2\beta\gamma}\frac{S}{\sqrt{A_c}} \tag{4-7}$$

$$\frac{1}{E_r} = \frac{1-\nu^2}{E} + \frac{1-\nu_i^2}{E_i} \tag{4-8}$$

$$\delta_c = \delta_{\max} - \varepsilon\frac{P_{\max}}{S} \tag{4-9}$$

where H is the hardness, $P_{\max}$ is the maximum indentation load, $A_c = C \cdot \delta_c^2$ is the projected area of tip-sample contact, C = constant of the area function (C = 24.56 for a perfect Berkovich indenter), δ_c is the contact depth, E_r is the reduced modulus, $S = \mathrm{d}P/\mathrm{d}\delta$ represents the experimentally measured stiffness (the slope of the unloading curve evaluated at the position of maximum load), P is the indenter load, δ is the indenter displacement, β is a dimensionless parameter related to the geometry of the indenter (β = 1.034 for a triangular punch), γ is a correction factor introduced by Hay et al. (1999) due to unrealistic boundary conditions used by Sneddon, E and E_i are the Young's moduli of sample and indenter, ν and ν_i are the Poisson's ratios of sample and indenter respectively, ε is a constant assumed to be 0.75.

This procedure can be successfully applied for materials with a high work hardening exponent which sink-in during indentation but for highly plastic materials like steels this procedure may cause significant errors due to the piling-up phenomenon (Y.-T Cheng & C.-M. Cheng, 2000). The conventional analysis method to extract hardness and modulus from nanoindentation data due to Oliver & Pharr (1992) tends to greatly overestimate the hardness and modulus due to the effect of "pile-up". In Figure 4-19 the material displaced by the Berkovich indenter pushes out to the sides of the indentation and forms a pile-up which supports some of the load, making the projected contact area larger than the cross-sectional area of the indenter at the original surface level.

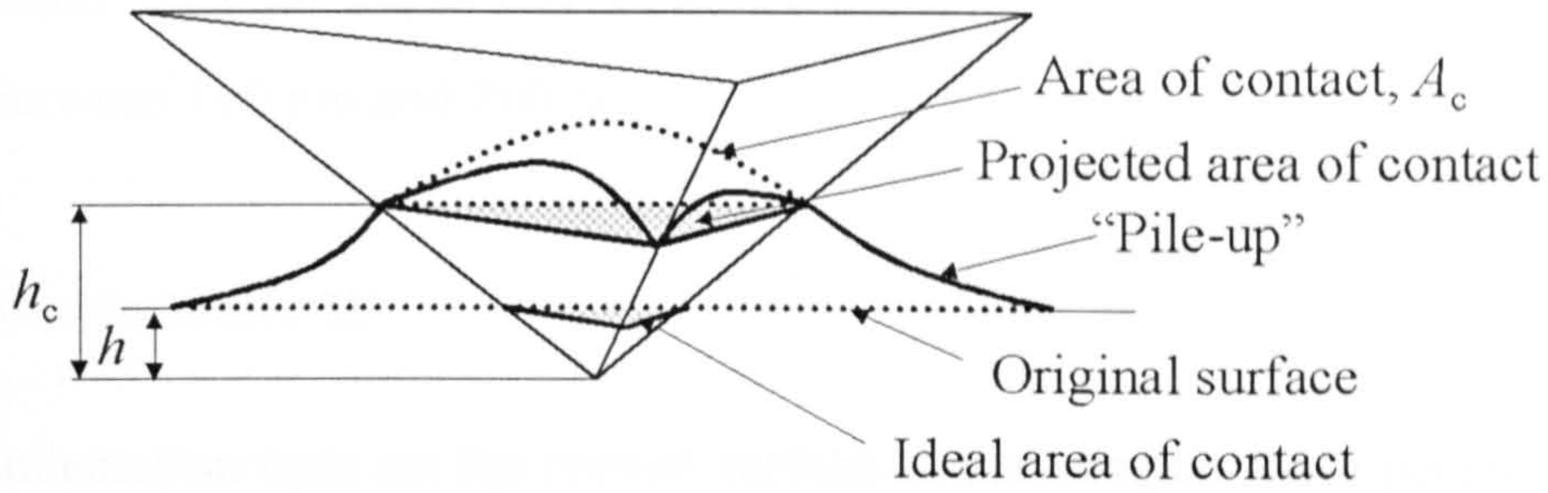


Figure 4-19. Pile-up effect in a Berkovich indenter/sample contact.

A second source of error in processing nanoindentation data relates to the deviation of the geometry of the indenter from its intended shape. Recently, Oliver (2001) suggested a new technique, called "the slope technique", for calculating hardness, modulus and contact area from indentation data, which substantially diminishes this effect but does not eliminate the effect of pile-up. The slope technique is based upon the equations 4-10 & 11.

$$H = \frac{1}{CP}\left(\frac{S_u S_l}{2S_u - \varepsilon S_l}\right)^2 \tag{4-10}$$

$$E_r = \sqrt{\frac{\pi}{C}}\frac{1}{2P\gamma}\left(\frac{S_u^2 S_l}{2S_u - \varepsilon S_l}\right) \tag{4-11}$$

where P is the indentation load, and S_l and S_u are the slopes of the loading and unloading curves, respectively. This approach has been previously shown to be less sensitive to the effects of pile-up (Bull, 2002). Oila & Bull (2003) have shown that the conventional analysis method due to Oliver and Pharr tends to overestimate the hardness and elastic modulus due to the effect of pile-up by about 25 %. Hardness and Young's modulus calculated using the slopes method are overestimated by about 10 %. They determined two correction factors for gear steels, $k_H = 0.82$ and $k_E = 0.75$ and applied in the Oliver and Pharr formulas for hardness and elastic modulus, respectively. The corrected values are in good agreement with those expected from the pile-up geometry.

Two types of nanoindenters have been used in this study: Nanoindenter II and Hysitron TriboIndenter. In both cases the tip used was the pyramidal Berkovich, which is a three-

sided pyramid with an angle 142.3° at its tip. The average radius of curvature is typically between 100 nm and 200 nm.

4.7.6.1 Nanoindenter II

The nanoindentation tests on the curved surface of the samples were performed with a Nanoindenter II™ manufactured by Nano Instruments, Knoxville, TN, USA, schematically shown in Figure 4-20.

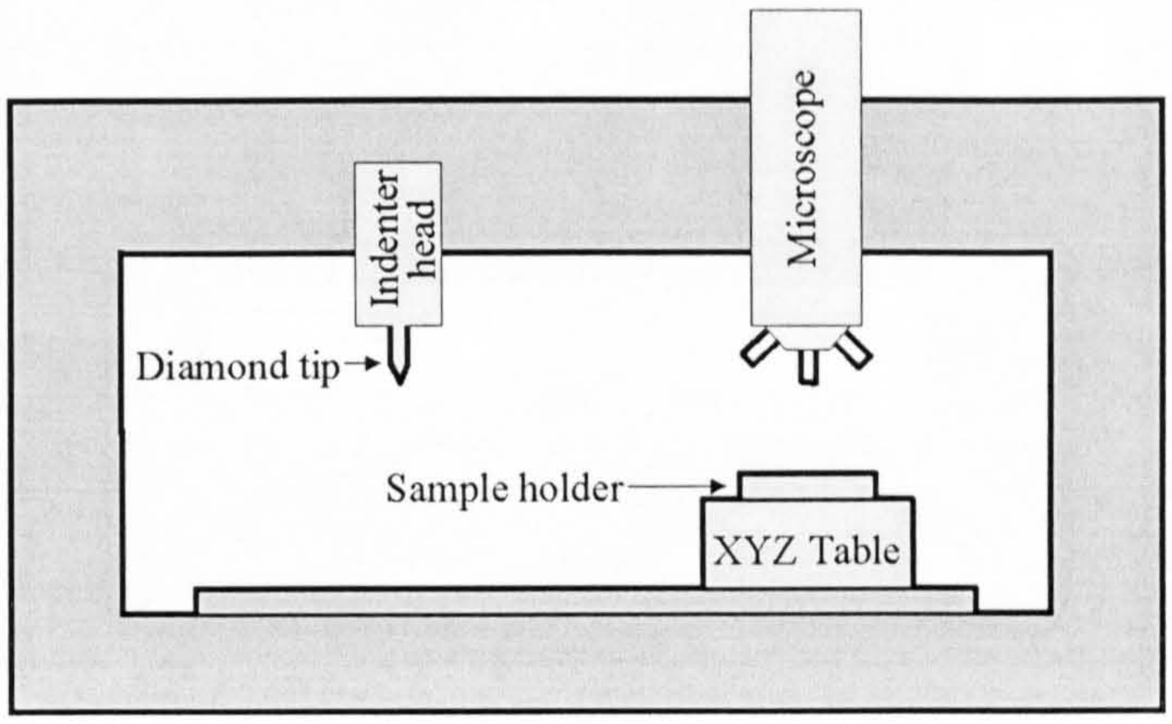


Figure 4-20. Nanoindenter II.

Loads ranging from 1 to 500 mN can be applied by means of an electromagnetic system, leading to indent depths ranging from a few nanometers to several micrometers. During the indentation test, load and displacement are recorded as a function of time.

4.7.6.2 Hysitron TriboIndenter

The Hysitron TriboIndenter recently acquired by University of Newcastle has been used to perform most of the nanoindentation tests, particularly those on polished cross sections. The TriboIndenter is a nanomechanical testing system designed to provide fully automated testing as well as in-situ imaging. The system is equipped with a TriboScanner which is designed to provide fine scale positioning of the indenter tip before and after conducting tests. The TriboScanner can also be used for imaging of the surface before and after indentation is performed. In-situ imaging is performed with the same tip used for indentation testing and allows pre and post-test observation of the

sample surface. This feature provides the capability to position the indenter probe within ten nanometers of the desired test location. The Hysitron's patented three-plate capacitive force/displacement transducer provides high sensitivity, large dynamic range, and a linear force or displacement output signal. The TriboScan 5.0 software controls all the functions of the TriboIndenter.

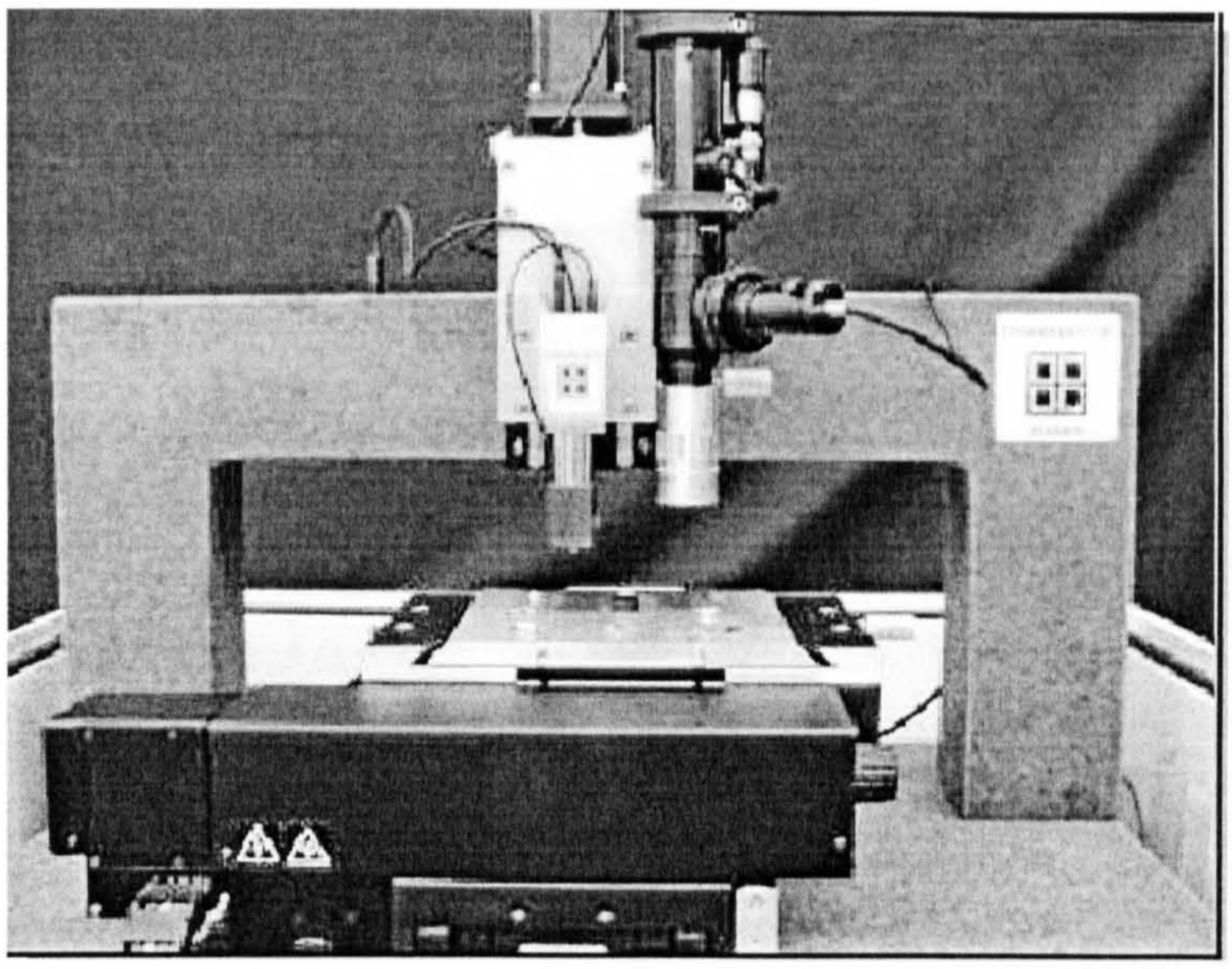


Figure 4-21. Hysitron TriboIndenter.

When imaging, the TriboIndenter performs as a scanning probe microscope (SPM) in a contact mode. The force between the tip and the sample is monitored using the capacitance sensor of the transducer. The voltage needed to move the piezo up and down, according to surface topography is monitored and correlated to the surface height.

The optics system with a CCD camera allows visual observation of the sample surface in the magnification range from x500 to x3500 and selection of testing locations. With the ability to resolve one-micron features, the tip can be placed within one micron of the desired surface feature. With these capabilities it is practically possible to test all phases that are present in the case of a carburised gear steel.

The TriboIndenter has been used to obtain hardness profiles from the longitudinal sections of the samples. The indentation pattern was a rectangular array of 10x10 indents with 20μm spacing between two consecutive indents. It was also used to test different microstructural features (inclusions, butterflies, white etching bands, etc).

4.7.7 X-ray diffraction

Changes in residual stress pattern and the transformation of retained austenite play a significant role in explaining the mechanism of micropitting and the decay of martensite (Voskamp et al., 1980).

Depending on their magnitude, residual stresses can be classified in *macrostresses*, often referred to as Type I and *microstresses* which are further classified as Type II and Type III (see Figure 4-22). Type I or macro residual stresses (σ') are residual stresses that develop on a scale larger than the grain size. Type II (σ'') are micro residual stresses that operate at the grain size level. Type III (σ''') are micro residual stresses generated at the atomic level. They exist, essentially as a result of the presence of dislocations.

Among the techniques used for measuring residual stresses (e.g., hole drilling, neutron diffraction, layer removal, Raman spectroscopy, etc), the X-ray diffraction technique (XRD) is ranked as second in popularity in a review of UK industrial survey carried out by Kandil et al. (2001). Only the XRD, neutron diffraction and Raman techniques have the necessary resolution to measure the microstresses.

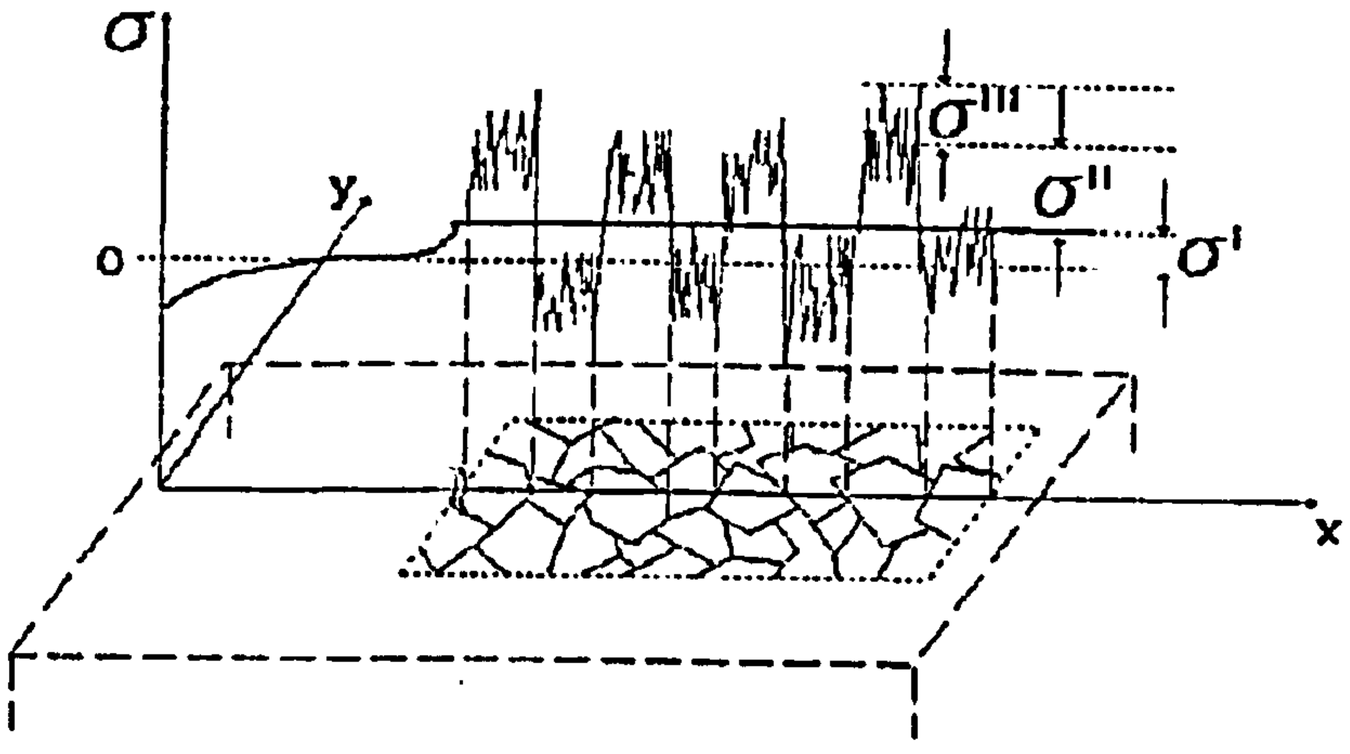


Figure 4-22. Types of residual stresses according to their magnitude (from Kandil et al., 2001)

Residual stress measurement techniques measure strains and the residual stresses are then deduced using the appropriate material parameters such as Young's modulus and Poisson's ratio. In the X-ray diffraction method the strain is measured in the crystal lattice, and the residual stress producing the strain is calculated assuming a linear elastic distortion of the crystal lattice (Prevey, 1996). The elastic deformations within a

polycrystalline material cause changes in the spacing of the lattice planes from their stress-free value to a new value that depends on the magnitude of the residual applied stress. This new spacing will be the same in any similarly oriented planes, with respect to the applied stress and the crystal lattice therefore acts as a very small strain gauge. During a measurement, the specimen is irradiated with high energy X-rays that penetrate the surface, the crystal planes diffract some of these X-rays, according to Bragg's law and a detector, which moves around the specimen to detect the angular positions where diffracted X-rays are located, records the intensity of these rays at that angular position. The locations of the peaks are used to evaluate the stress within the specimen.

Several methods can be used to evaluate the stresses using X-ray diffraction technique. The most popular method is $\sin^2\Psi$ method. In the $\sin^2\Psi$ method an interplanar spacing serves as an internal strain gauge. If $d_{\Phi\Psi}$ is the spacing between the lattice planes measured in the direction defined by ψ and Φ, the strain can be expressed as

$$\varepsilon_{\Phi\Psi} = \frac{d_{\Phi\Psi} - d_0}{d_0} \tag{4-12}$$

where d_0 is the stress-free lattice spacing. Measurements are made at different values of Ψ and the residual stress is calculated from the d-$\sin^2\Psi$ plot.

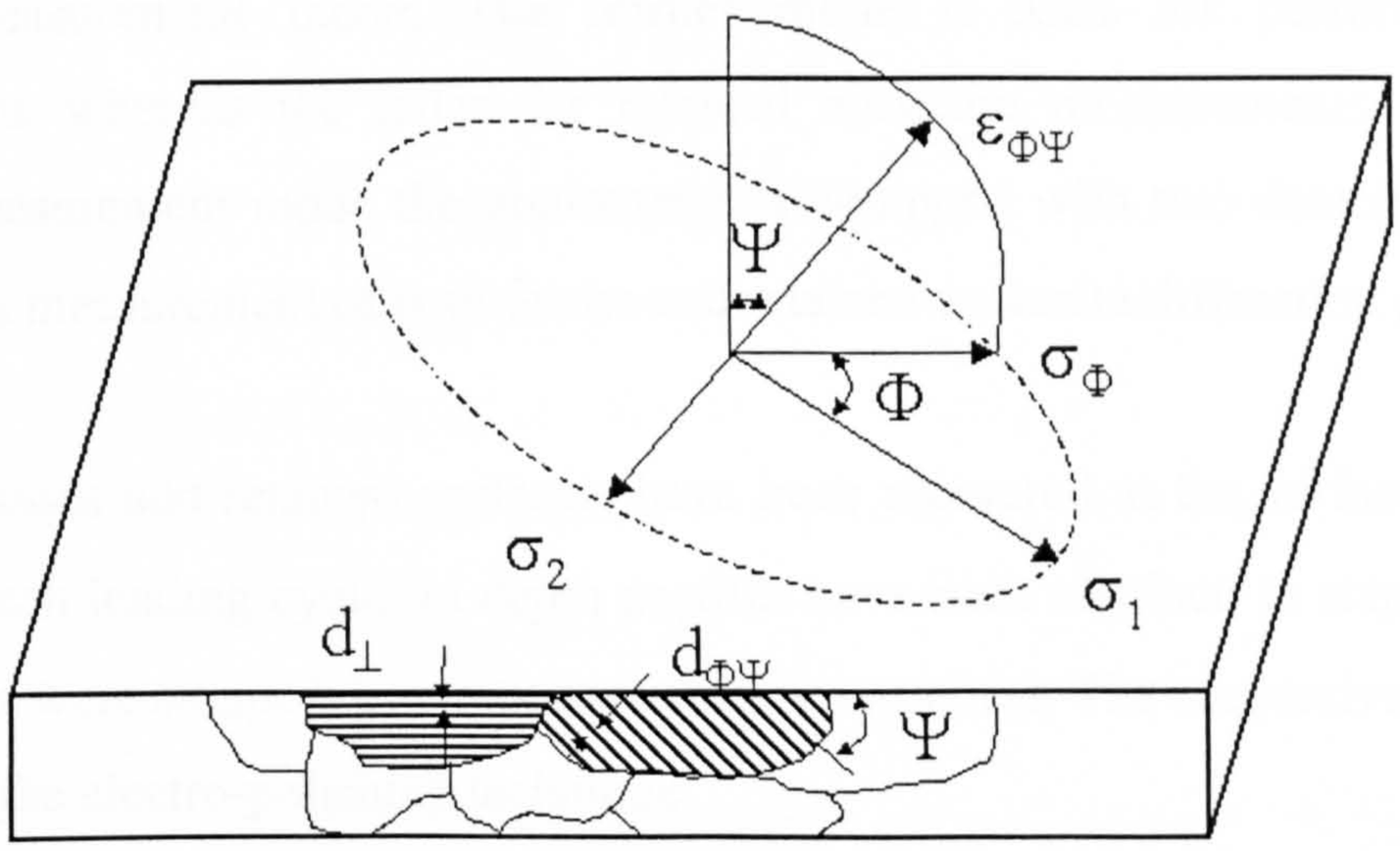


Figure 4-23. Definition of the angles ψ and Φ.

Retained austenite content can be estimated using metallographic techniques down to concentrations on the order of 15 vol. %. At lower concentrations metallographic

analysis becomes difficult to apply. X-ray diffraction provides the method of choice. It is also more accurate and with detection limits to 0.2 vol. %.

Residual stresses and retained austenite have been measured in this work using a XSTRESS 3000 X-ray stress analyser produced by Stresstech Group. The XSTRESS 3000 consists of a central unit and a goniometer (see Figure 4-24).

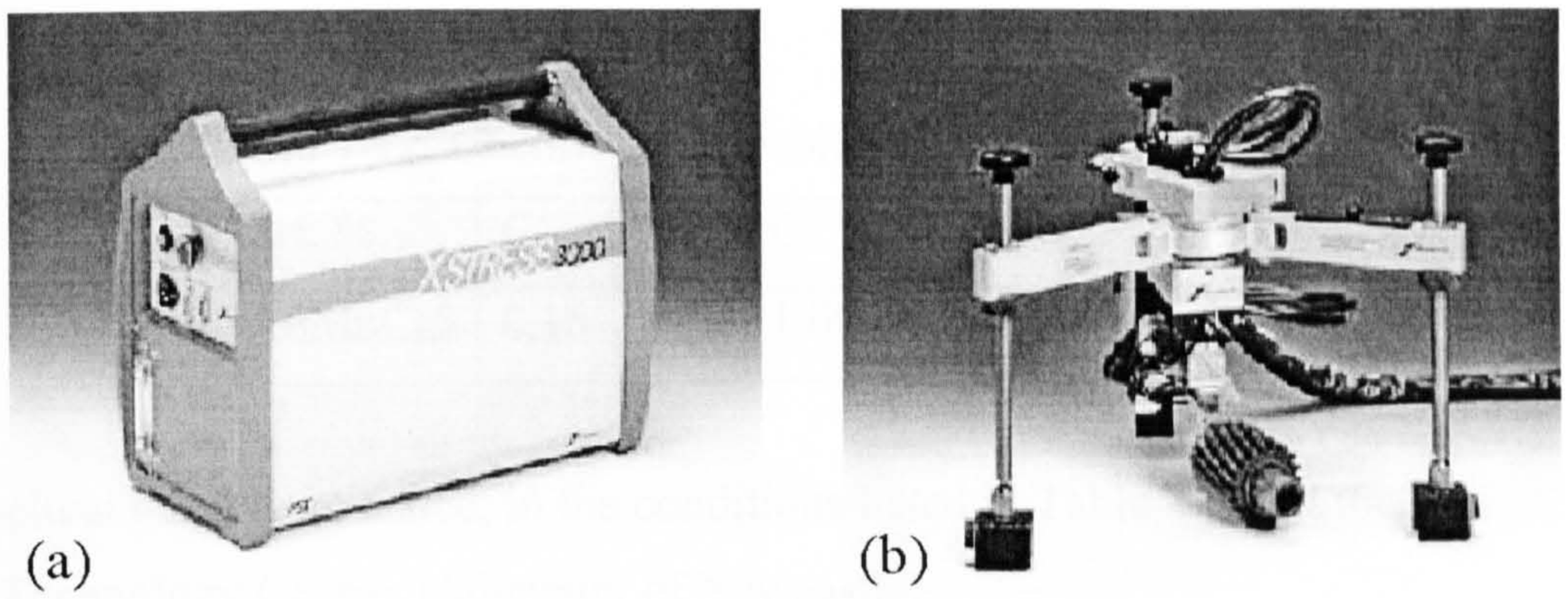


Figure 4-24. XSTRESS 3000 X-ray stress analyser (a) central unit and (b) goniometer.

The central unit contains the X-ray power supply and recirculating water cooling with heat exchanger for X-ray tube and power supply. The goniometer is equipped with a Cr X-ray tube and replaceable collimators of variable diameters (1, 2, 3, 4, and 5 mm spot size). The stress analyser can operate in d-sin$^2\Psi$ measurement mode and retained austenite measurement mode. The former mode is used for performing stress measurements whereas the latter for retained austenite measurements. In retained austenite measurement mode the goniometer is equipped with two detectors allowing simultaneous measurement of both ferrite and retained austenite diffraction peaks.

Residual stresses and retained austenite have been measured at the surface of the test discs after each loading cycle. In depth profiles have been obtained in steps of 10 μm, after the tests were stopped, by successive material removal. The successive layers were removed by the electro-polishing technique.

4.7.8 Investigation into gears

In order to confirm the micropitting results obtained from the twin disc machine, similar investigations have been carried out on carburised gears. The gears analysed in this study were helical gears made from 16MnCr5 steel. The chemical composition of the steel is given in Table 4-12.

Table 4-12. Chemical composition of the 16MnCr5 steel.

wt. %	C	Mn	Cr	Si	S	P
16MnCr5	0.16	1.15	1.00	0.30	0.035	0.035

The helical gears were tested, in the conditions listed in Table 4-13, at the Design Unit - Gear Technology Centre, University of Newcastle.

Table 4-13. Operating conditions in the Design Unit gear test.

Gear code	Lubricant	Temperature	Load	Speed
B12	OEP-80	90	1547	3000
B25	OEP-80	50	1547	3000
B28	OEP-80	90	1547	3000
B29	OEP-80	50	1547	3000
B34	OEP-80	90	1547	3000
C15	OEP-80	50	1547	1125
C79	OEP-80	90	1547	1125
C34	OEP-80	50	1300	3000
C44	OEP-80	50	1700	1125
C62	OEP-80	50	1700	1125
C76	OEP-80	50	1300	3000

All the investigations reported in this work were performed post-test. The topography of the surface affected by micropitting has been studied using light microscopy and optical profilometry on xantopren replicas taken from the gear tooth surface after the first stage of testing.

Residual stresses have been measured at the pitch line and on locations where surface damage has taken place.

In order to perform the metallographic examination the gear tooth was sectioned in transverse direction, perpendicular to the direction of the surface lay. In order to ensure the edge retention during grinding and polishing, gear tooth samples were nickel plated using the same technique as for disc specimens. The samples were then mounted in carbon filled epoxy resin in order to avoid the occurrence of the charging effect during SEM examination. After they were mounted the samples were ground successively on 240 and 600 SiC papers lubricated by water and then polished on alumina cloths of 9, 3 and 1 μm size. The polished specimens were used to observe cracks and non-metallic inclusions. The microstructure was observed after etching with 2 % nital. In order to determine the mechanical properties of different phases present in the gear steel the nanoindentation testing was carried out using the Hysitron triboindenter.

4.8 Summary

Chapter 4 describes the experimental techniques used to investigate micropitting in disc specimens and gears. It also gives brief descriptions of the equipments used to assess micropitting. The results obtained from the micropitting tests are presented in the next chapters.

~ Chapter 5 ~

MICROPITTING EVALUATION

5 MICROPITTING EVALUATION

5.1 General

This chapter presents the results of the investigations carried out during tests on discs. The results refer to the driven disc and they characterise the surface topography, the morphology of micropitting and microcracks and, the micropitting initiation and propagation rate. These results are correlated with hardness, residual stress and retained austenite measurements. In order to evaluate stresses due to asperity contact, the measured surface roughness data were applied in the Greenwood-Williamson asperity contact model. The rate of wear, which has been evaluated from the mass losses recorded after each testing stage, indicates that the loss of material is caused by micropitting. No damage caused by wear debris was observed; hence no further investigations have been carried out on this matter. Because the radii of curvature of the specimens change after the first testing stage, two sets of elastohydrodynamic parameters have been calculated to describe the initial and the final conditions for each test. The number and the size of the micropits have been accurately determined by the method described in section 4.7.4. Microscopic examination of the microcracks confirmed that micropitting is a phenomenon driven by crack initiation and propagation.

5.2 Surface topography

The topography of the tested surfaces has been investigated using light and scanning electron microscopy and, optical profilometry. The evolution of micropitting, during testing, can be followed on the images taken with the light microscope after each testing stage. Figure 5-1 shows the appearance of the disc surface, at different testing stages, in experiment 2. The direction of the contact motion is O*y*. It can be seen that micropitting extends as the number of cycles increases and the grinding marks tend to flatten. The micropitting extends by the coalescence of the existing micropits as well as by the

initiation of new micropits. As cycling continues, the micropits grow and merge with their neighbours. At the same time, new micropits are formed on the remaining grinding marks, which now support the load.

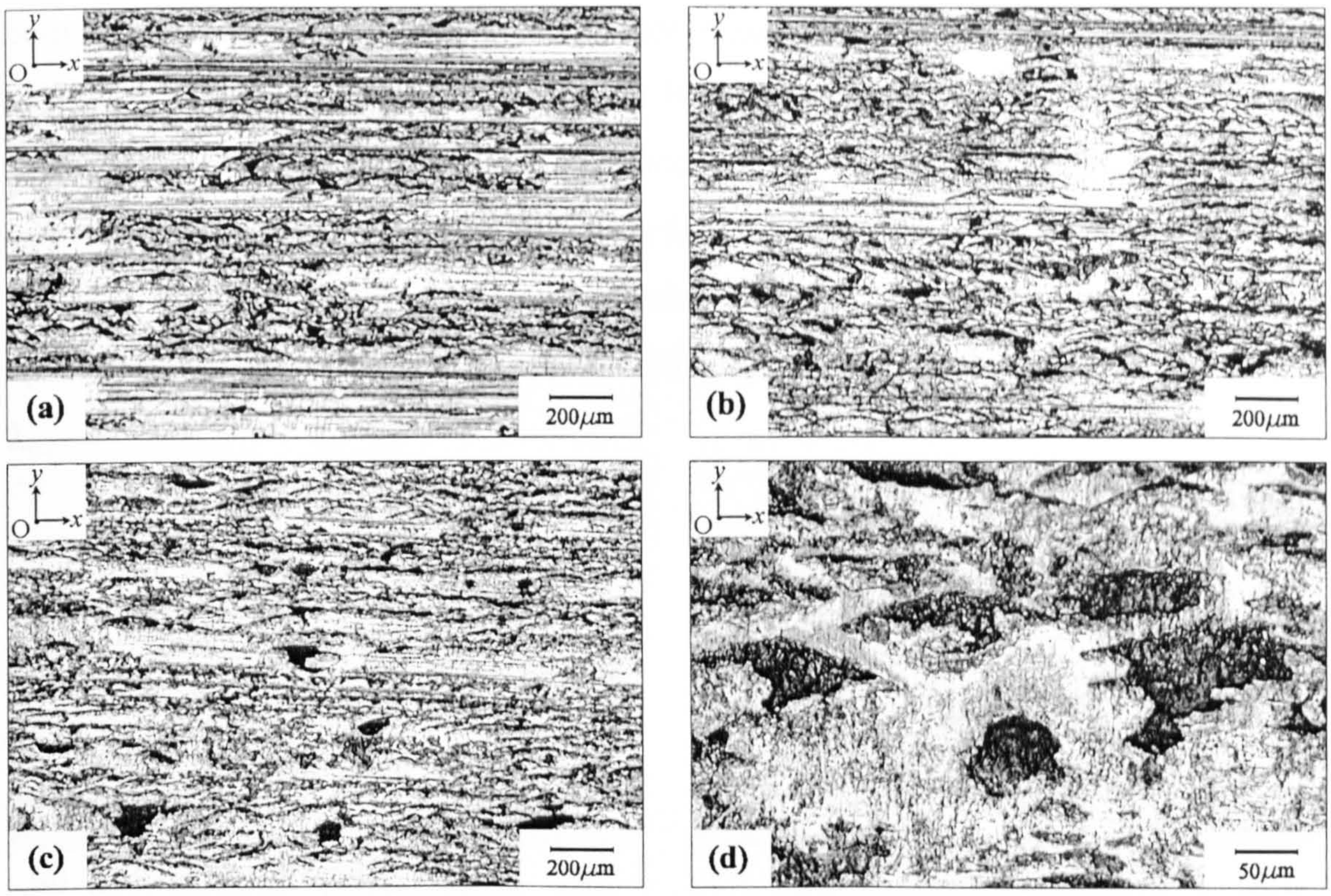


Figure 5-1. Surface appearance of the sample in experiment 2, after: (a) $1.8 \cdot 10^5$ cycles (b) $3.6 \cdot 10^5$ cycles, (c) and (d) $5.4 \cdot 10^5$ cycles. Sliding direction is O*y*.

The shape of the micropits (see Figure 5-1d) is elongated in the direction of the surface lay (O*x*). This is explained by the geometry of the asperities, which have a larger length in this given direction. Figure 5-2 shows that the micropits are initiated on the top of the highest marks left on the surface by the grinding process.

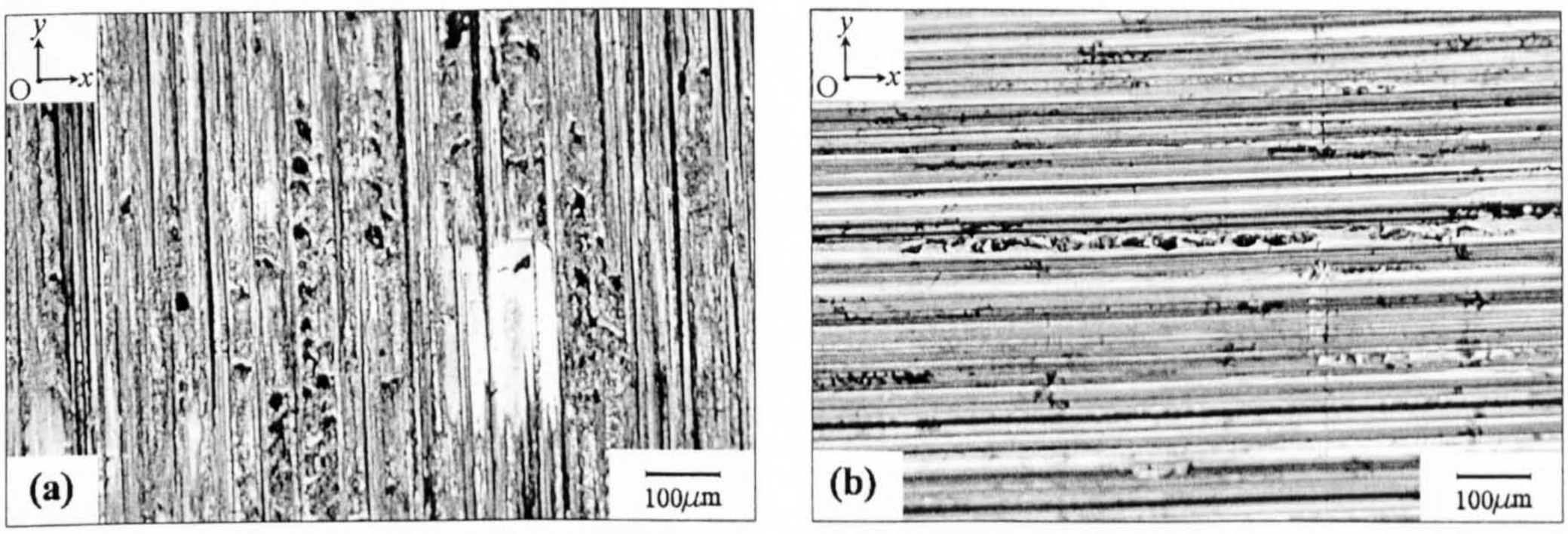


Figure 5-2. Initiation of micropitting: (a) experiment 1 ($3.6 \cdot 10^5$ cycles) and (b) experiment 3 ($5.4 \cdot 10^5$ cycles). Sliding direction is O*y*.

The specimens from Figure 5-2 were subjected to different grinding processes, which resulted in longitudinal surface lay in (a) and transverse surface lay in (b). The micropits are aligned in the direction of the surface lay, which is parallel to the direction of the contact motion (O*y*) in Figure 5-2a, and perpendicular in Figure 5-2b. Their occurrence follows a certain periodicity, which can be approximated by the wavelength of the surface profile in the direction of the surface lay, $\lambda_{q//}$. In Figure 5-3 the micropits appear in parallel bands that are parallel to the surface lay. This suggests a periodic distribution of the micropits in the direction perpendicular to the surface lay, which can be estimated by the wavelength of the surface profile in the direction perpendicular to the surface lay, $\lambda_{q\perp}$.

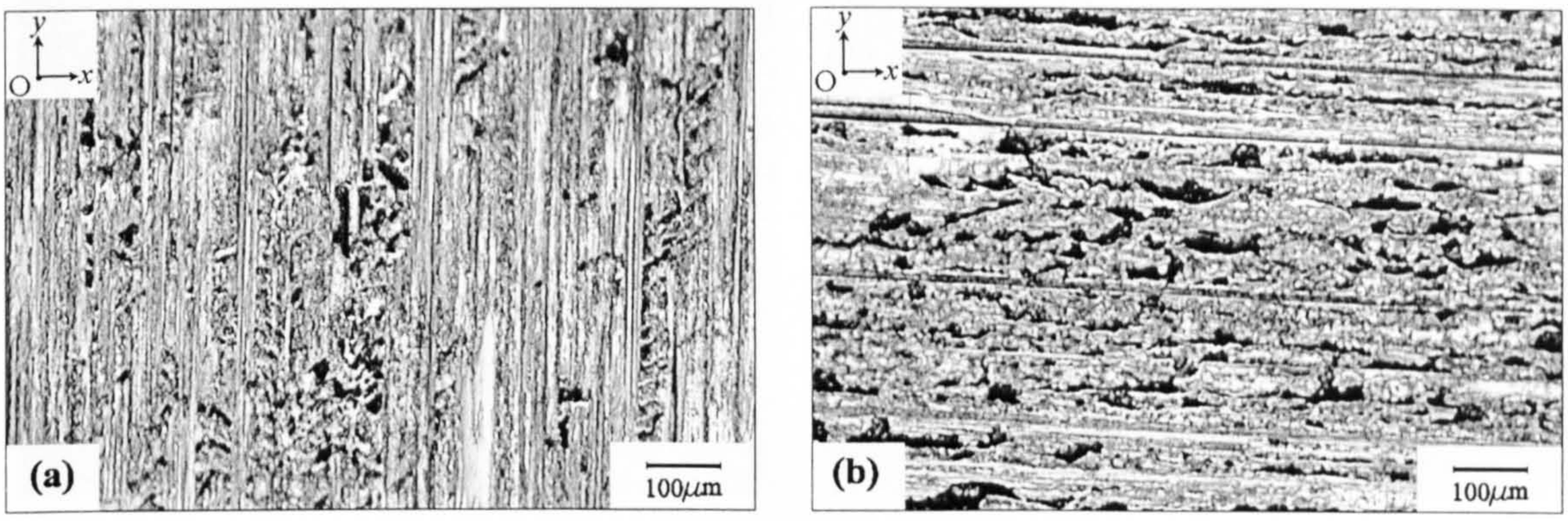


Figure 5-3. Surface appearance of the sample in (a) experiment 1 ($3.6 \cdot 10^5$ cycles), (b) experiment 6 ($1.8 \cdot 10^5$ cycles). Sliding direction is O*y*.

SEM observation of the affected surface revealed that the asperities have been subjected to severe plastic deformation. This is well illustrated in Figure 5-4a in which a number of asperities appear to be detaching from the surface.

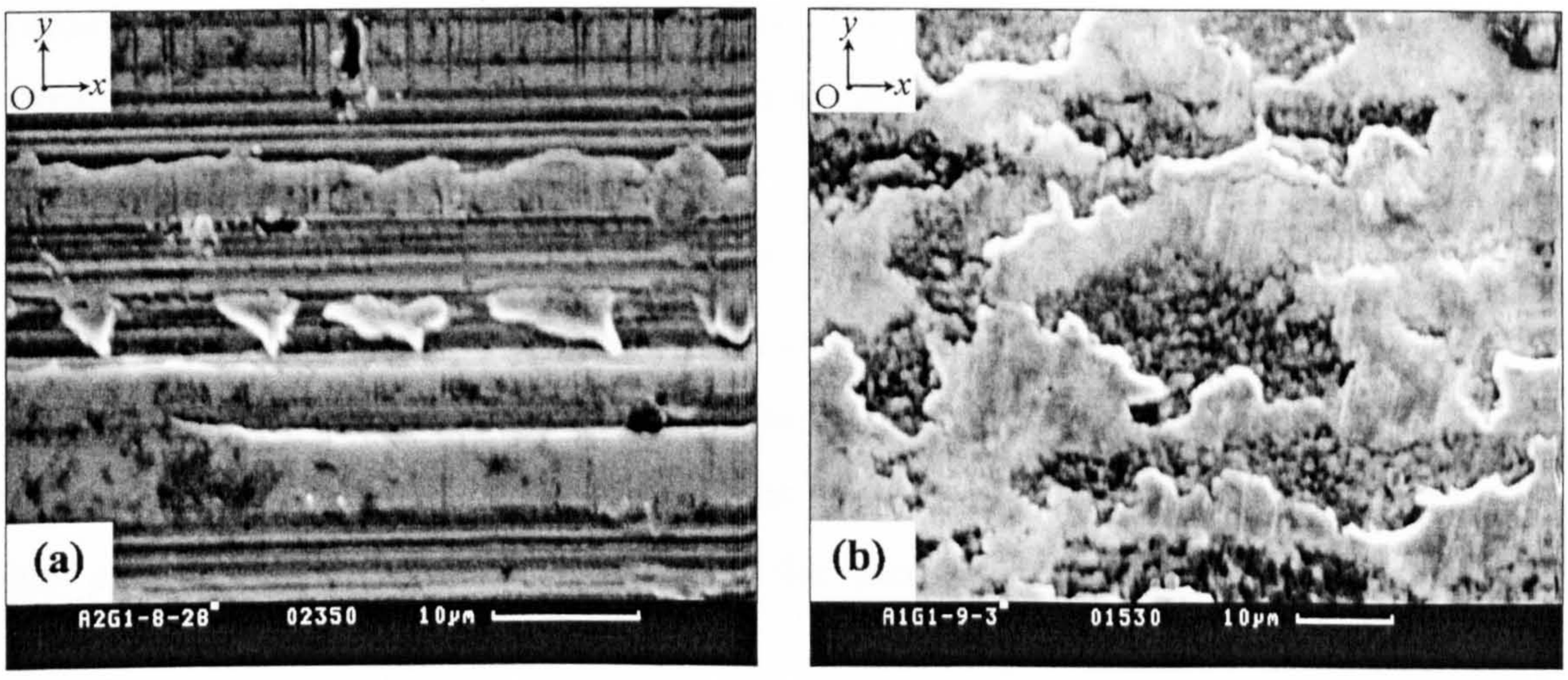


Figure 5-4. SEM images of the affected surface. (a) plastic deformation of asperities (experiment 3), (b) micropits (experiment 6). Sliding direction is O*y*.

In Figure 5-4b it can be clearly seen that material at the edges of the micropit is plastically deformed in the O*y* direction which is the sliding direction. The rough appearance of the surface inside a micropit (see Figure 5-5), which resembles a surface resulting from ductile fracture, also suggests that the material has undergone plastic deformation before fracture. The micropits shown in Figure 5-5 confirm that they are elongated in the direction of the surface lay, i.e., O*y* in 5-5a and O*x* in 5-5b. This implies that the asperities have an ellipsoidal geometry with the maximum semi-axis oriented in the direction of the surface lay.

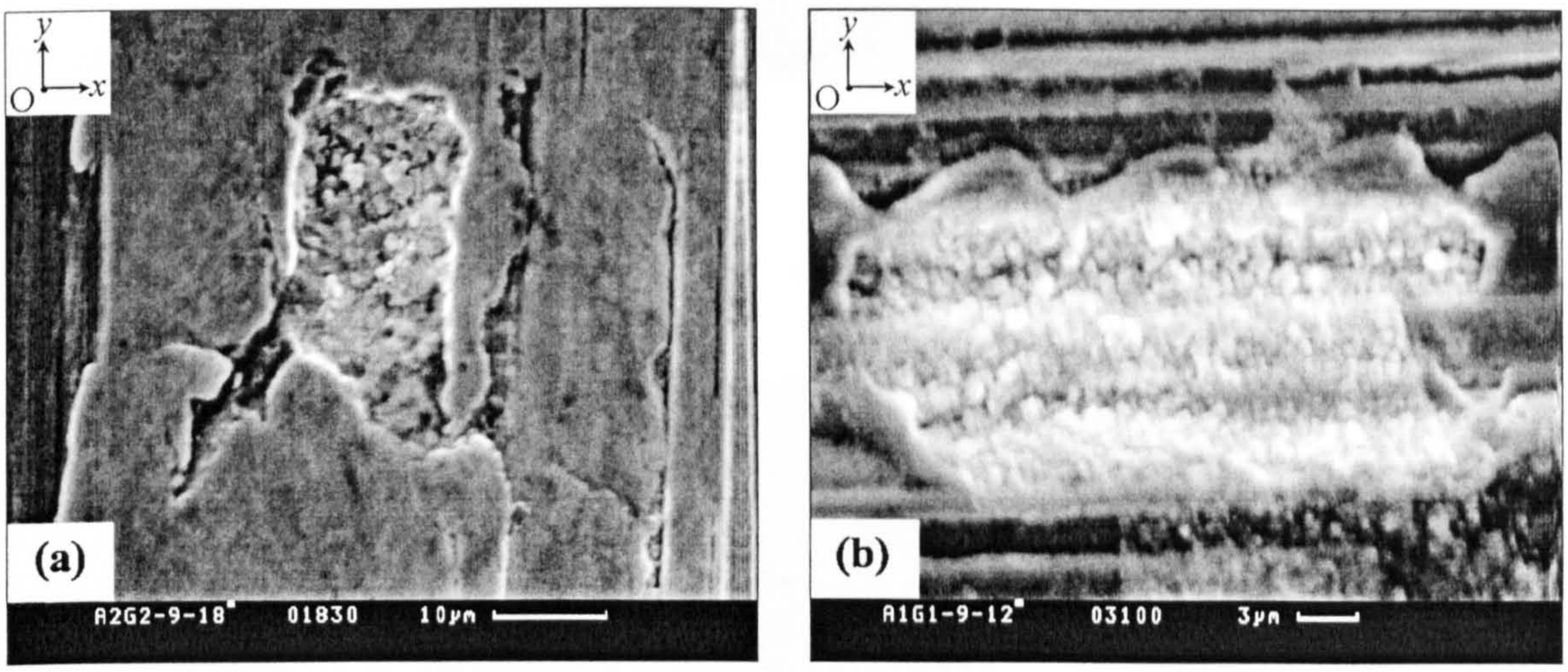


Figure 5-5. The appearance of the fracture surface. (a) experiment 7 ($3.6 \cdot 10^5$ cycles) and (b) experiment 6 ($1.8 \cdot 10^5$ cycles). Sliding direction is O*y*.

The optical profilometer allows for the observation of the surface in three dimensions. It can be clearly seen in Figures 5-6 and 5-7 that the deformation of asperities and the formation of micropits are linked together. The figures show that the micropits are surrounded by a pile-up of material, which suggests that they have been formed where an asperity was present. In Figure 5-6b the region marked by the white square is further magnified and rotated by 180° to allow a better observation of the details. The aforementioned observation concerning the rough appearance of the fracture surface can also be noted in Figure 5-6b. Although the surface in the vicinity of the micropit is smooth, it is likely that the remaining peaks bordering the micropits will contribute, by their plastic deformation during cycling, to the micropit coalescence.

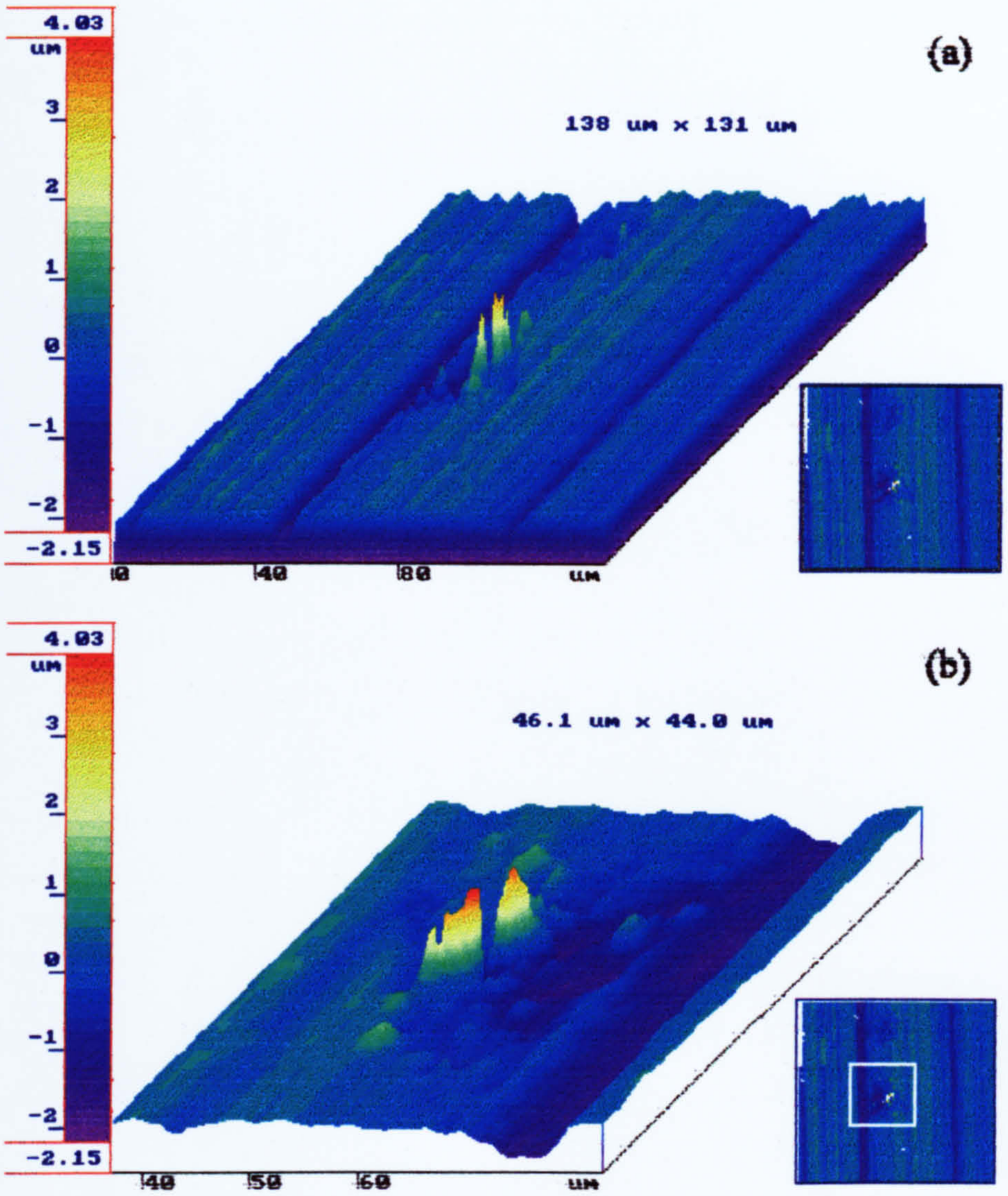


Figure 5-6. Optical profilometer image (Experiment 3, $1.8 \cdot 10^5$ cycles). (a) micropit bordered by remnant asperity; (b) higher magnification of the framed area.

Compared to Figure 5-6, the specimen shown in Figure 5-7 has been subjected to higher contact stresses and, as a consequence, the plastic deformation was more severe (the surface appears smoother and the grinding marks are not distinguishable anymore) and, as a result, more micropits have been formed. Figure 5-7b shows a higher magnification of the region marked by the black square. The figure shows an undeformed asperity next to a micropit. This suggests that the height of this asperity was smaller than the height of the asperity where the micropit has formed and it was subjected to a lower stress or, perhaps, it did not take part to the contact at all.

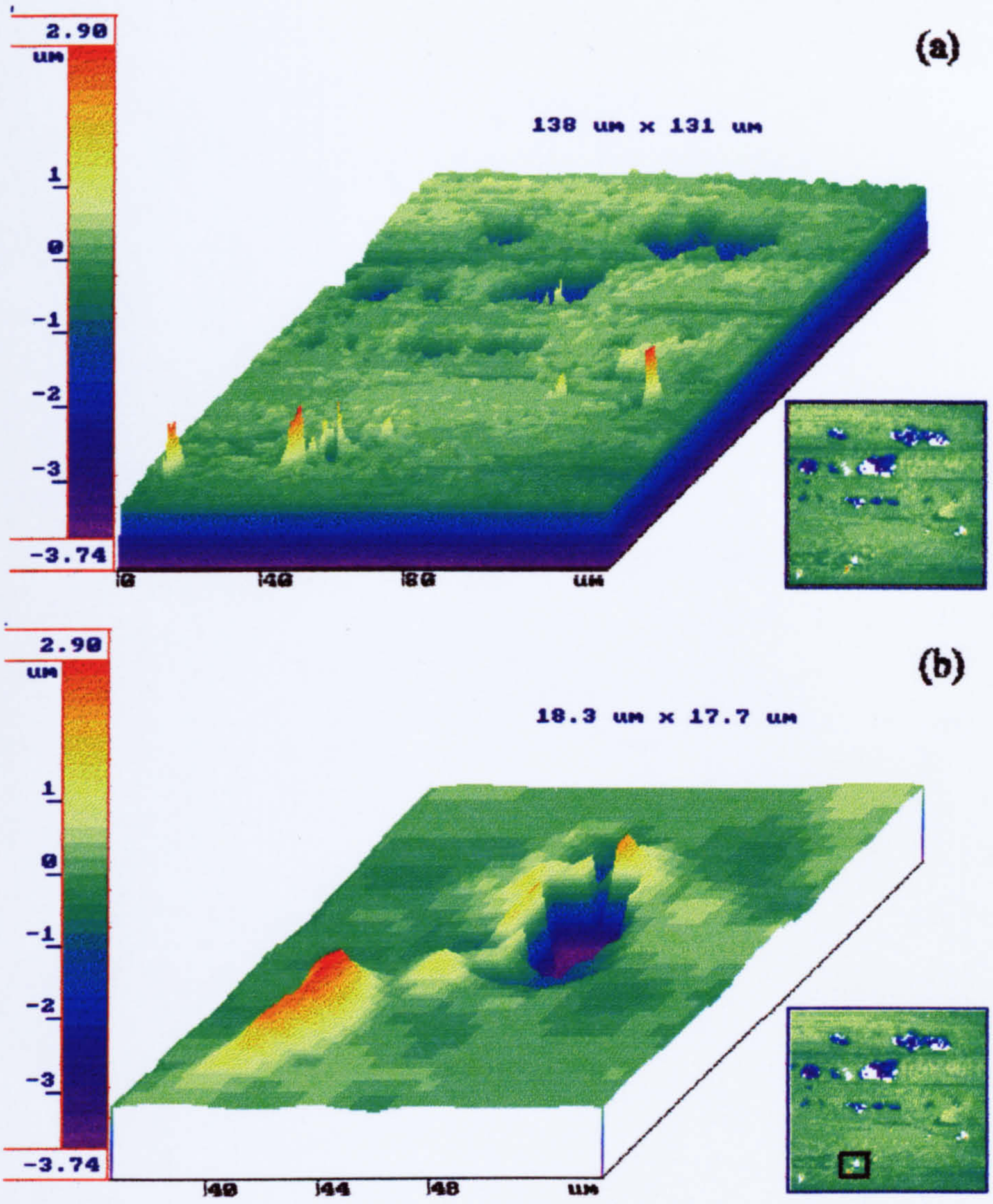


Figure 5-7. Optical profilometer image (experiment 7, $3.6 \cdot 10^5$ cycles). (a) micropits and remnant asperities; (b) higher magnification of the framed area.

5.3 Wear rate

The formation of micropits and consequently material removal leads to mass loss of the disc specimen. The recorded mass losses in each experiment are listed in Table 5-1. These data have been used to plot the cumulative mass loss against the number of stress cycles in Figure 5-8. The mass loss curves have been fitted to lines in Figure 5-9. The slopes of these lines (i.e., the ratio of mass loss to the number of cycles, dm/dN) represent the *mass loss rates* or the *wear rates* specific to each experiment. The wear rates are plotted as columns in Figure 5-10. As a function of the magnitude of the wear

rate, the experiments can be classified in three levels, as follows: experiments 6 and 1-high wear rate, experiment 2-moderate wear, experiment 3,4,5,7 and, 8-low wear rate.

Table 5-1. Mass losses recorded in the eight experiments.

Number of cycles, N	Mass loss, $\Delta m \cdot 10^{-3}$ (mg)							
	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8
$1.8 \cdot 10^5$	21.400	4.806	3.204	1.197	1.403	21.294	4.196	2.304
$3.6 \cdot 10^5$	17.899	12.39	0.793	1.602	5.798	-	2.197	3.295
$5.4 \cdot 10^5$	-	3.204	1.602	3.196	-	-	-	4.600
$7.2 \cdot 10^5$	-	-	11.99	-	-	-	-	4.600

In experiment 6 all the factors were at their high level, which explains the high wear rate. However, experiment 1 was carried out at low load but high speed and slide-to-roll ratio and the recorded wear rate is still high. This implies that increasing speed and slide-to-roll ratio result in increased wear rate. In experiment 3 all the factors were at their low level, which explains the lowest wear rate. However, in experiment 4 the temperature and slide-to-roll ratio were at the high level but the wear rate is low.

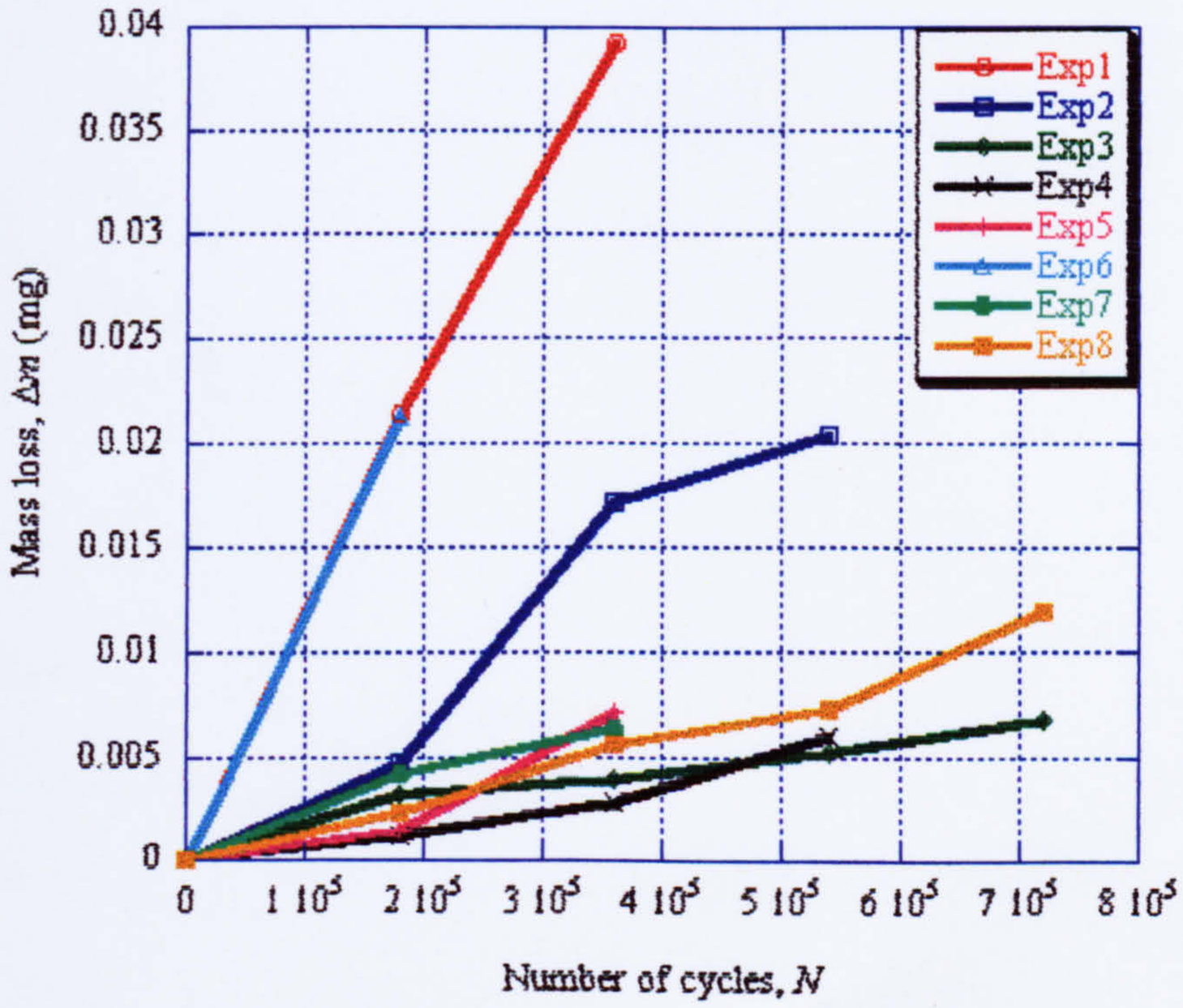


Figure 5-8. Mass loss as a function of number of cycles.

It is difficult to evaluate the influence of seven factors only by using these plots. A more elaborated mathematical approach is needed. These considerations justify the use of the factorial experimental design in order to assess the factors of influence.

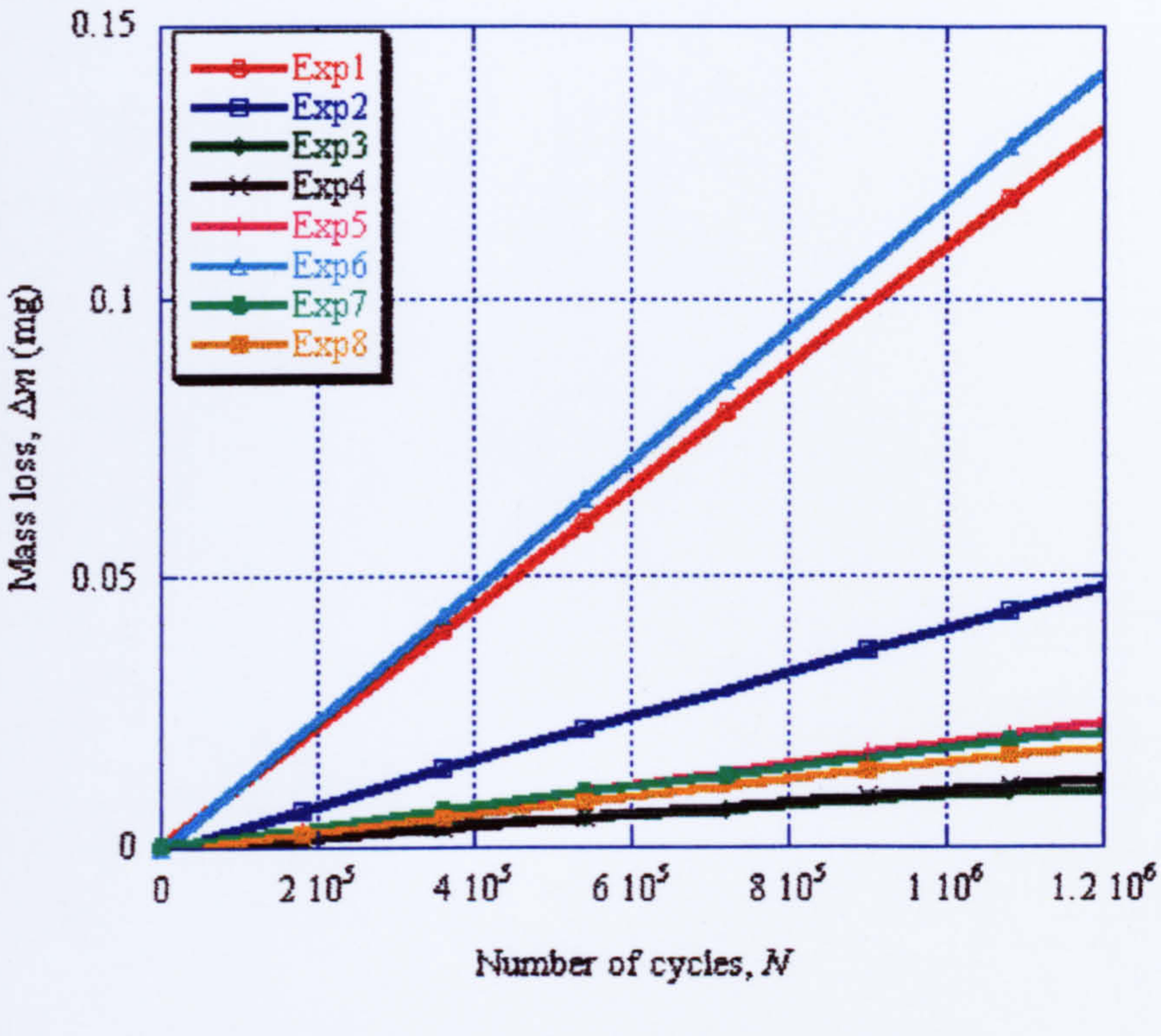


Figure 5-9. Mass loss curves fitted by lines.

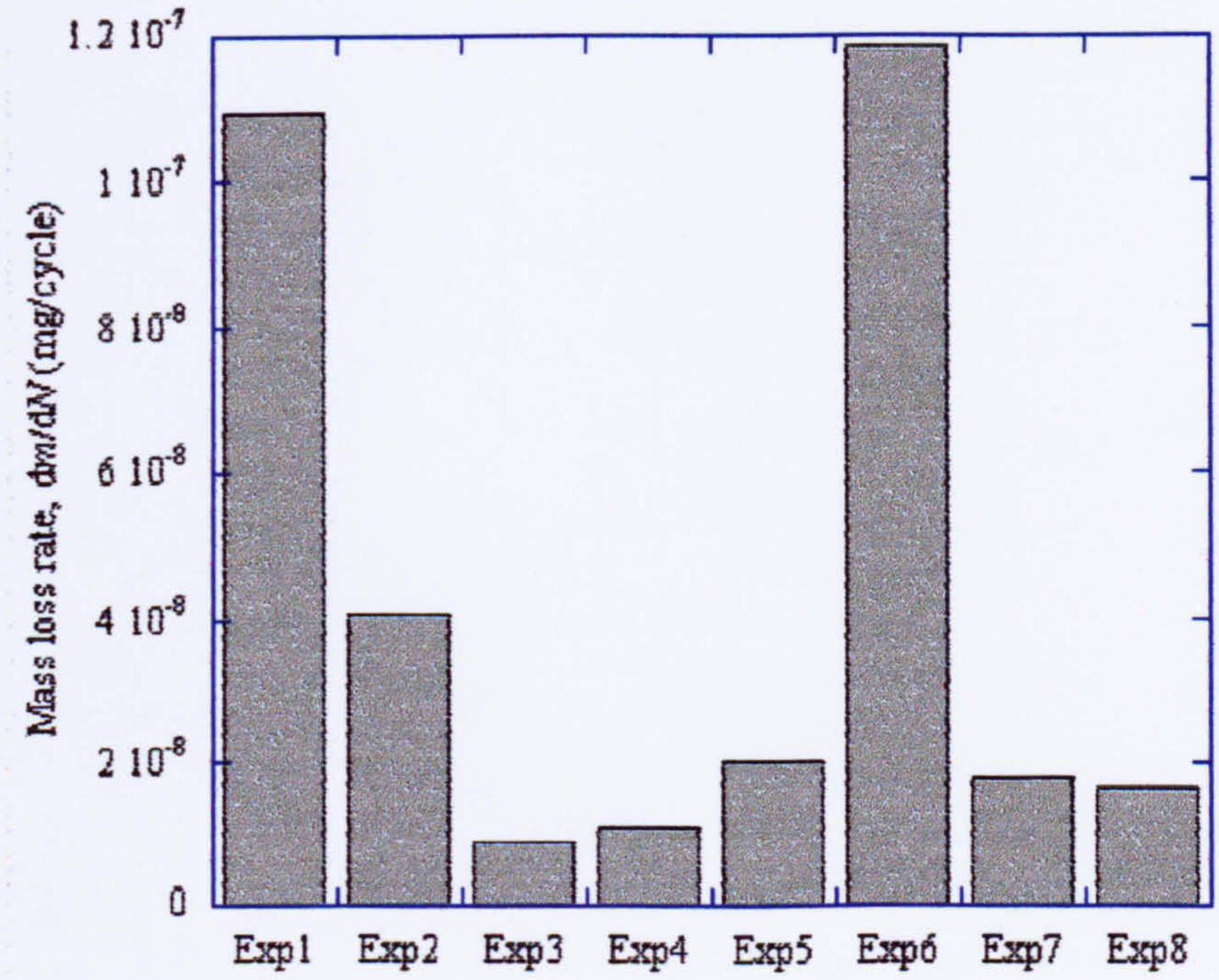


Figure 5-10. Mass loss rate, d*m*/d*N*.

5.4 Changes in the operating conditions

The loss of material, as well as the plastic deformation of asperities, leads to changes of the radii of curvature of the disc specimens, which ultimately leads to changes in the operating conditions. Only the maximum radii of curvature, R_1' and R_2', perpendicular to the direction of contact motion vary significantly. The subscripts 1 and 2 denote the driven and the driving disc, respectively. These changes were recorded (see Table 5-2) after the first testing stage ($1.8 \cdot 10^5$ cycles) but no changes were detected at a later stage. Therefore, the radii of curvature measured before testing were used in the calculations of the elastohydrodynamic parameters corresponding to the *initial contact conditions* and the radii of curvature measured after the first testing stage were used to calculate the parameters corresponding to the *final contact conditions*.

Table 5-2. The radii of curvature of the disc samples. Initial and final conditions.

Experiment		R_1'(mm)	R_1''(mm)	R_2'(mm)	R_2''(mm)	R'(mm)	R''(mm)	R_e(mm)
1	initial	124.97	24.98	132.25	34.97	64.25	14.57	30.59
	final	144.28	24.975	295.60	34.95	96.95	14.56	37.58
2	initial	128.67	26.48	144.70	33.495	68.10	14.78	31.73
	final	257.23	26.465	289.32	33.48	136.1	14.78	44.86
3	initial	105.08	26.49	144.41	33.36	60.82	14.76	29.96
	final	181.91	26.47	204.19	33.35	96.20	14.75	37.67
4	initial	111.87	25.02	147.90	34.99	63.69	14.58	30.48
	final	250.01	25.00	209.12	34.98	113.8	14.57	40.74
5	initial	115.13	26.50	129.44	33.5	60.93	14.79	30.02
	final	181.96	26.485	289.32	33.48	111.7	14.78	40.64
6	initial	144.27	24.97	111.78	34.97	62.98	14.56	30.29
	final	249.81	24.96	147.82	34.955	92.87	14.56	36.77
7	initial	148.55	26.475	118.13	33.48	65.80	14.78	31.19
	final	257.23	26.465	129.39	33.475	86.09	14.78	35.67
8	initial	144.35	25.00	132.33	35.01	69.04	14.58	31.73
	final	249.96	24.99	295.77	34.99	135.4	14.57	44.44

Table 5-3. Initial and final operating conditions in the eight micropitting experiments.

	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8
p_0 (GPa)	1.8	2.2	1.8	1.8	1.8	2.2	2.2	2.2
	1.5	2.0	1.5	1.5	1.5	2.0	2.0	2.0
a (μm)	1330	1697	1291	1306	1274	1660	1692	1713
	1634	2400	1624	1746	1725	2015	1936	2400
b (μm)	494	613	502	488	496	625	625	607
	461	546	465	443	448	586	598	543
δ (μm)	21	32	21	20	21	34	33	32
	20	29	20	19	19	32	32	29
z (μm)	329	407	334	325	329	415	415	404
	344	407	309	330	333	436	445	404
τ_{max} (MPa)	585	740	593	603	609	742	728	740
	505	582	509	492	493	646	659	586
u (m/s)	3.76	3.76	3.13	3.13	3.76	3.76	3.13	3.13
	3.76	3.76	3.13	3.13	3.76	3.76	3.13	3.13
G	2361	2463	2506	2298	2266	2172	2203	2614
	2361	2463	2506	2298	2266	2172	2203	2614
$U \cdot 10^{-11}$	5.42	5.01	6.36	3.05	3.47	3.59	3.11	5.78
	4.41	3.54	5.06	2.28	2.56	2.95	2.72	4.12
$W \cdot 10^{-5}$	2.48	4.43	2.59	2.40	2.47	5.07	4.78	4.43
	1.64	2.21	1.64	1.34	1.35	3.44	3.66	2.26
α (GPa^{-1})	22.13	22.13	23.49	20.65	20.36	20.36	20.65	23.49
η (Pas)	$9.3 \cdot 10^{15}$	$6.5 \cdot 10^{19}$	$1.5 \cdot 10^{17}$	$4.5 \cdot 10^{14}$	$2.4 \cdot 10^{14}$	$8.5 \cdot 10^{17}$	$1.7 \cdot 10^{18}$	$1.8 \cdot 10^{21}$
	$1.2 \cdot 10^{13}$	$9.3 \cdot 10^{15}$	$1.3 \cdot 10^{14}$	$9.3 \cdot 10^{11}$	$5.5 \cdot 10^{11}$	$1.4 \cdot 10^{16}$	$2.8 \cdot 10^{16}$	$1.5 \cdot 10^{17}$
h_{cen} (μm)	0.72	0.69	0.82	0.48	0.52	0.50	0.47	0.79
	0.75	0.74	0.85	0.51	0.55	0.51	0.48	0.84
h_{min} (μm)	0.25	0.23	0.29	0.16	0.18	0.17	0.16	0.26
	0.21	0.18	0.25	0.13	0.14	0.15	0.14	0.20

The initial and final contact conditions for each experiment are listed in Table 5-3. The parameters taken into consideration are as follows: the maximum contact pressure, p_0, the semi-axis of the contact ellipse, a and b, the approach of centres, δ, the depth of maximum shear stress, z, the maximum shear stress, τ_{max}, the entraining velocity, u, the elastohydrodynamic non-dimensional parameters, G, U, and W, the pressure-viscosity coefficient, α, the lubricant viscosity, η, and the central and minimum lubricant film

thickness, h_{cen} and h_{min}. The data given in Table 5-3 show that, after the initial stages, the area of contact increases and, consequently, the contact pressure decreases. The increase of the contact area is consistent with the observations made in the previous sections that the surface tend to flatten due to plastic deformation of asperities. The depth of maximum shear stress lies between 300 and 450 μm. This is much higher than the observed depth of micropits, i.e., in the order of microns, therefore it can be concluded that Hertzian macro-stresses are not related to micropitting.

5.5 Surface roughness measurements

The Hertzian theory alone does not describe the contact conditions with sufficient accuracy since it does not account for the roughness of the surface. As the microscopic examination of the surface indicates (section 5.2) the surface conditions change continuously during test. At the beginning the highest asperities undergo severe plastic deformation, which leads to the formation of micropits. As a consequence, new asperities come into contact and the load will be redistributed on these contacting asperities having the highest height. As cycling continues, new micropits are being formed, the existing micropits grow and the height of the grinding marks reduce until the entire surface is flattened. One useful means to monitor these changes is represented by the surface roughness measurement after each testing cycle. The surface roughness parameters determined by the optical profilometer are listed in Table 5-4. The data represent the mean of five measurements taken in five different locations randomly chosen on the wear track. The micropits are uniformly distributed on the wear track.

As mentioned in section 5-2, the surface roughness wavelength has a significant influence on micropitting initiation and propagation. The root mean square wavelength, λ_q, in two orthogonal directions has been calculated using formula 2-40 and the data from Table 5-4. Since, after the first testing stage, the values of the profile slopes in the directions Ox and Oy, Δ_{qx} and Δ_{qy}, do not vary significantly, these values were averaged in order to obtain the final value for the profile slopes. The same procedure has been applied for root mean square height, R_q. The ratio of the wavelength in the direction perpendicular to the direction of the contact motion, λ_{qy} to the wavelength in the direction of the contact motion, λ_{qx} can be used to characterise the surface roughness.

Table 5-4. Surface roughness parameters.

Exp	$N \cdot 10^5$	R_q (μm)	R_a (μm)	R_p (μm)	R_v (μm)	R_t (μm)	Δ_{qx} (mrad)	Δ_{qy} (mrad)
Exp1	0	0.46	0.37	1.61	-1.60	3.21	0.13	0.34
	1.8	0.38	0.28	2.10	-2.34	4.44	0.23	0.34
	3.6	0.46	0.35	2.11	-2.79	4.90	0.32	0.42
Exp2	0	0.25	0.19	0.75	-1.17	1.92	0.17	0.08
	1.8	0.37	0.26	0.81	-2.70	3.51	0.19	0.15
	3.6	0.37	0.24	0.81	-2.88	3.69	0.20	0.14
	5.4	0.34	0.23	0.79	-2.85	3.64	0.18	0.15
Exp3	0	0.60	0.48	2.51	-1.86	4.38	0.39	0.18
	1.8	0.34	0.27	1.05	-1.56	2.61	0.12	0.05
	3.6	0.20	0.16	0.70	-1.15	1.85	0.11	0.05
	5.4	0.46	0.36	1.89	-2.21	4.11	0.41	0.21
	7.2	0.35	0.27	0.73	-1.97	2.69	0.17	0.06
Exp4	0	0.33	0.26	1.18	-1.32	2.50	0.21	0.09
	1.8	0.29	0.20	0.80	-1.93	2.73	0.17	0.11
	3.6	0.43	0.30	0.81	-2.99	3.80	0.23	0.13
	5.4	0.45	0.32	0.79	-3.31	4.09	0.20	0.13
Exp5	0	0.47	0.38	1.16	-1.52	2.68	0.07	0.21
	1.8	0.29	0.22	0.70	-2.01	2.71	0.09	0.16
	3.6	0.31	0.22	1.14	-2.82	3.97	0.12	0.17
Exp6	0	0.85	0.71	2.17	-2.12	4.29	0.32	0.09
	1.8	0.63	0.43	1.55	-4.54	6.09	0.25	0.16
Exp7	0	0.29	0.24	0.80	-1.02	1.82	0.03	0.13
	1.8	0.23	0.16	0.78	-1.86	2.64	0.04	0.07
	3.6	0.48	0.30	0.80	-3.89	4.69	0.09	0.12
Exp8	0	0.38	0.30	1.29	-1.53	2.82	0.09	0.22
	1.8	0.21	0.15	0.60	-1.04	1.64	0.07	0.14
	3.6	0.20	0.14	0.94	-1.28	2.22	0.07	0.12
	5.4	0.18	0.14	0.50	-1.29	1.79	0.07	0.13
	7.2	0.32	0.21	0.86	-2.54	3.40	0.09	0.15

The wavelengths in the two directions, λ_{qx} and λ_{qy}, the wavelength ratio $\Lambda = \lambda_{qy}/\lambda_{qx}$, the composite roughness R_q* and the specific film thickness (or lamda ratio), λ, specific to

each experiment are given in Table 5-5. Figure 5-11 shows the surface profiles of the specimen tested in experiment 1 in the direction perpendicular to the surface lay in (a) and perpendicular in (b). The wavelength in the perpendicular direction is approximately twice the wavelength in the parallel direction.

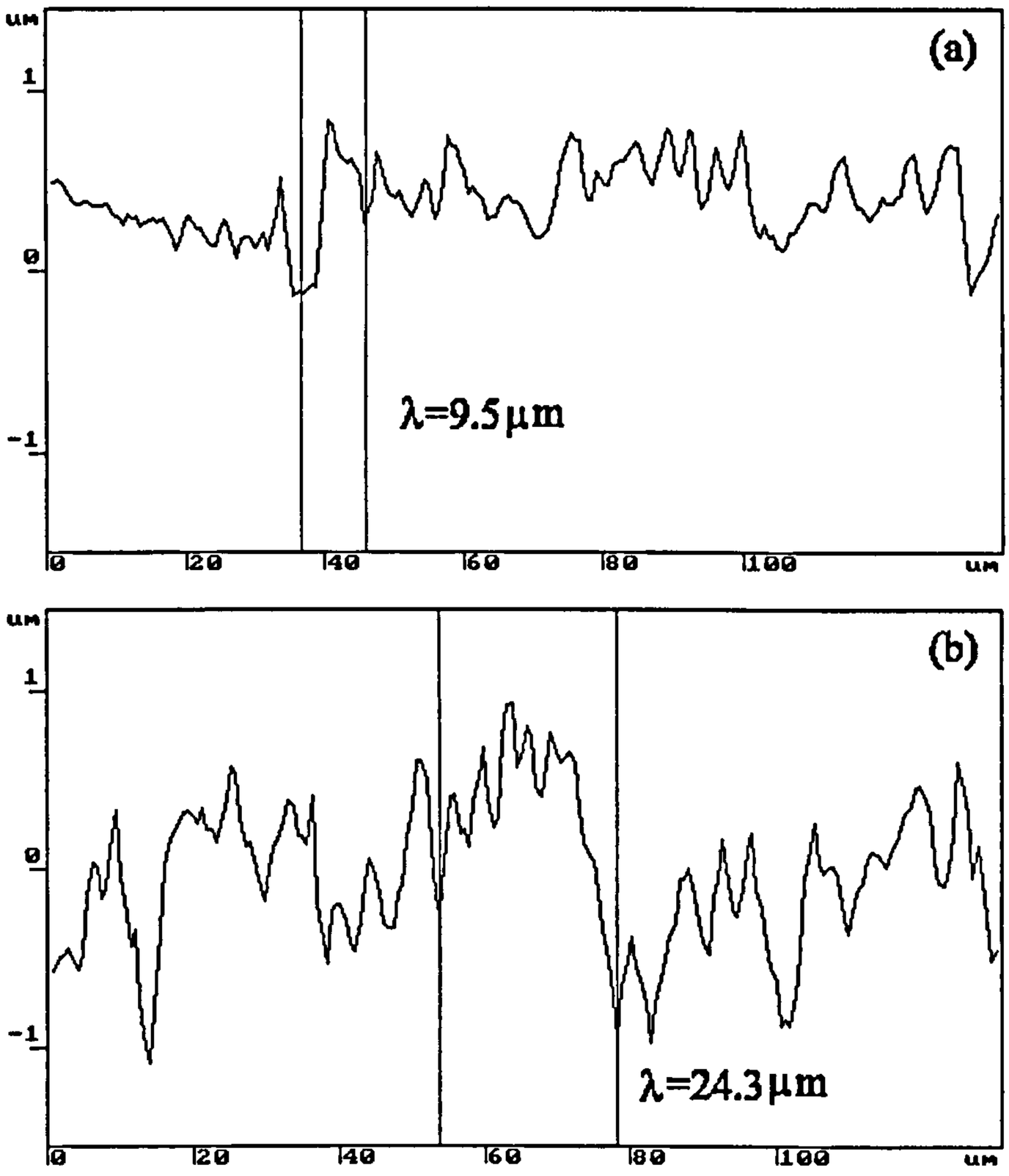


Figure 5-11. Surface profile wavelength (a) parallel to surface lay; (b) perpendicular.

Due to the proprietary rights, the rheological properties of the lubricants are unknown therefore any attempt to calculate the friction coefficient would be unsuccessful. The friction coefficient within lubricant has been calculated by Olver & Spikes (1998) for a two-discs test. They found a value $\mu_f = 0.05$ and showed that the friction coefficient does not vary significantly for different contact pressures. Using this value and the formula 2-97, the friction coefficient specific to the tests carried out in this work has been calculated and the results are given in Table 5-5.

The specific film thickness lies in the range 0.1 to 1 and the friction coefficient from 0.08 to 0.13. These values indicate that the contact conditions correspond to a regime of

mixed lubrication (see Table 2-2) in which the metal-to-metal contact plays a significant role.

Table 5-5. Wavelengths, composite roughness, lambda ratio and friction coefficient.

Experiment		λ_{qx} (μm)	λ_{qy} (μm)	$\Lambda=\lambda_{qy}/\lambda_{qx}$	R_q (μm)	λ	μ
1	initial	22.48	8.62	0.38	0.65	0.38	0.102
	final	9.84	6.97	0.71	0.59	0.35	0.104
2	initial	9.07	20.00	2.21	0.35	0.66	0.086
	final	11.94	15.81	1.32	0.51	0.35	0.104
3	initial	9.58	20.95	2.19	0.85	0.34	0.105
	final	12.23	28.52	2.28	0.47	0.57	0.090
4	initial	10.12	23.15	2.29	0.47	0.34	0.105
	final	12.25	19.53	1.59	0.55	0.24	0.115
5	initial	43.73	13.82	0.32	0.67	0.27	0.112
	final	18.67	11.36	0.61	0.42	0.33	0.106
6	initial	16.70	58.37	3.50	1.21	0.14	0.126
	final	16.16	24.11	1.49	0.89	0.17	0.123
7	initial	64.39	14.34	0.22	0.42	0.38	0.102
	final	32.03	22.72	0.69	0.50	0.31	0.108
8	initial	26.27	10.95	0.42	0.54	0.48	0.095
	final	18.56	10.45	0.56	0.31	0.65	0.086

The surface roughness parameters listed in Table 5-4 are not sufficient to investigate micropitting since micropitting is related to asperity contacts. It follows that the calculation of asperity parameters is necessary. The asperity parameters defined by the equations 2-54, 56, 58 and 60 in the Greenwood-Williamson model are: the summit density, μ_s, the mean summit height, r_s, the standard deviation of the summit heights, σ_s and the mean summit curvature, β^*. These parameters have been calculated for two orthogonal directions, in the direction parallel to the surface lay and in the direction perpendicular to the surface lay and the results suggests that the asperities have an ellipsoidal shape rather than a spherical one. This is consistent with the general observation that micropits are elongated in the direction of the surface lay. In order to be

able to apply the Greenwood-Williamson model, in which the asperities are approximated with spherical caps, the values obtained in two orthogonal directions were averaged. These data are given in Table 5-6. The root mean square radius of curvature, σ_k, has been calculated using equation 2-54 in which the root mean square slope, Δ_q, is a value from Table 5-4 and the summit density, μ_s, has been calculated as the ratio of the number of summits to the area of contact, A_c, calculated by the Hertzian theory, as below.

$$\mu_s = \frac{N_s}{A_c} \tag{5-1}$$

The number of summits, N_s, has been determined as the product of the number of peaks in Ox direction, N_{px}, and Oy direction, N_{py}.

$$N_s = N_{px} \cdot N_{py} \tag{5-2}$$

Considering that one summit is defined by two zero crossings of the profile, the number of peaks have been calculated as the ratio of the semi-axis of the Hertzian contact to the profile wavelength.

$$N_{px} = \frac{b}{\lambda_{qx}} \text{ and } N_{py} = \frac{a}{\lambda_{qy}} \tag{5-3}$$

It is now possible to apply the Hertzian theory to the contact between two hemispherical asperities that are characterised by the parameters given in Table 5-6. The contact between asperities is characterised by the radius of contact area a_i, asperity normal load, P_i, the normal approach of interference δ_i, asperity maximum contact pressure, p_{i0}, depth of maximum shear stress below asperity contact, z_i, and the maximum shear stress at the depth z_i, $\tau_{\iota max}$. Most of the previous researches (Shotter, 1980) have shown that the area of asperity contacts represents about 10 % from the total area of contact. The Abbott-Firestone curves (see Figure 5-12) suggest similar values. The parameters given in Table 5-7 have been calculated with the assumption that 10 % of the total area of contact is represented by the asperity contacts.

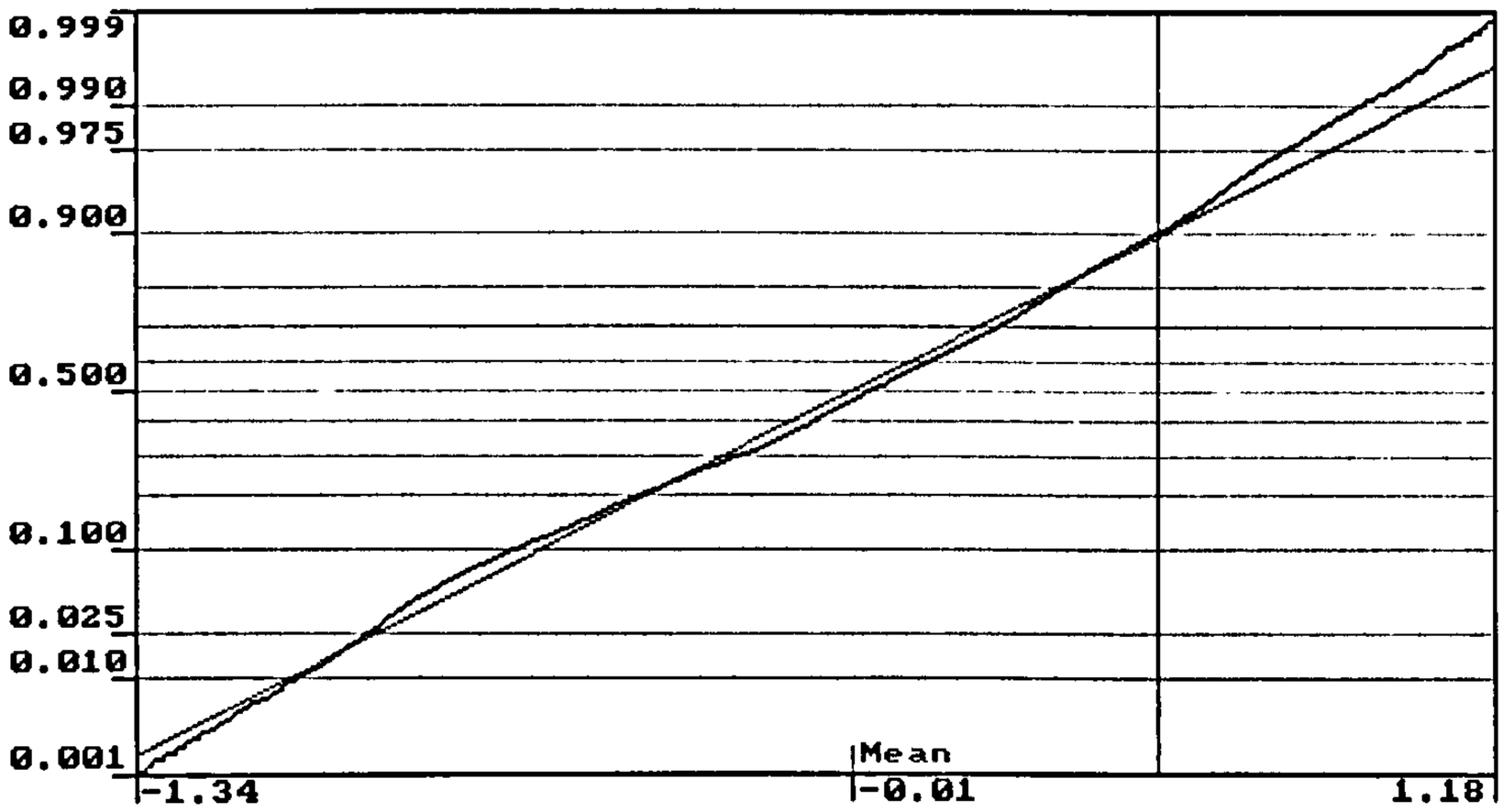


Figure 5-12. Bearing area curve. Experiment 1 before test. Peaks represent 10% of the profile.

Table 5-6. Asperity parameters.

Exp		A_c (mm^2)	μ_s (mm^{-2})	σ_k (μm)	r_s (μm)	β*(μm)	σ_s (μm)
1	initial	2.06	1642	53.72	2.26	80.83	0.21
	final	2.37	4668	127.7	1.87	192.1	0.17
2	initial	3.27	1755	29.83	1.17	44.88	0.06
	final	4.12	1693	39.28	1.60	59.10	0.13
3	initial	2.04	1585	65.18	2.84	98.07	0.36
	final	2.37	1393	36.55	1.61	54.99	0.12
4	initial	2.00	1358	31.11	1.58	46.80	0.11
	final	2.43	1381	33.86	1.76	50.94	0.15
5	initial	1.99	526	18.55	2.43	27.9	0.22
	final	2.43	1518	29.88	1.35	44.95	0.09
6	initial	3.26	326	21.35	4.51	32.13	0.73
	final	3.71	816	33.57	2.84	50.51	0.40
7	initial	3.32	344	8.38	1.68	12.61	0.09
	final	3.64	453	9.93	1.57	14.95	0.14
8	initial	3.27	1105	29.34	1.83	44.13	0.14
	final	4.09	1744	24.57	1.03	36.97	0.05

Table 5-7. Hertzian theory applied to asperities.

Experiment		a_i (μm)	P_i (N)	δ_i (μm)	p_{i0} (GPa)	z_i (μm)	τ_{imax} (GPa)
1	initial	14.92	5.85	2.76	12.54	7.16	3.89
	final	28.81	17.94	4.39	10.43	13.83	3.23
2	initial	24.43	48.20	13.3	38.56	11.73	11.95
	final	30.20	69.27	15.43	36.19	14.49	11.22
3	initial	14.46	4.39	2.13	10.01	6.94	3.10
	final	14.66	11.13	5.54	29.68	7.04	9.20
4	initial	13.17	7.24	3.70	19.93	6.32	6.18
	final	16.01	12.40	5.10	22.34	7.68	6.92
5	initial	8.13	2.86	2.37	20.64	3.90	6.40
	final	16.85	15.87	6.36	26.87	8.08	8.33
6	initial	10.51	5.14	3.44	22.21	5.04	6.89
	final	18.92	19.06	7.09	25.43	9.08	7.88
7	initial	11.01	15.04	9.61	59.27	5.28	18.37
	final	13.78	27.72	13.85	67.04	6.61	20.78
8	initial	19.38	24.48	8.51	31.11	9.30	9.64
	final	30.22	114.1	24.82	57.72	14.50	17.89

The depth of maximum shear stress, below asperity contact, lies between 3 and 15 μm. Since the depth of the micropits is in a similar range it follows that micropitting phenomenon is related to asperity contact rather than the bulk Hertzian contact.

5.6 Micropitting measurements

In order to assess the influence of the factors involved in micropitting development it is essential to accurately determine the severity of the surface distress. In the present work, the micropits are sufficiently uniformly distributed on the wear track, so that the method described in section 4.7.4 can be reliably used for micropitting measurements. Using the Scion Image software it is possible to count the number of micropits, to determine the area of the individual micropits and the total area occupied by the micropits. The

measurements were done on images obtained with the optical profilometer from five different locations situated on the wear track. The images used in these measurements are the same used for the measurement of the surface roughness parameters. The total area on which the measurements were taken is 138x131=18078 μm^2.

We define a *micropitting parameter*, *M*, as in the equation given below

$$M(\%) = \frac{\text{micropitted area}}{18078} \cdot 100 \tag{5-4}$$

Table 5-8. Micropitting measurements.

Experiment	No. cycles	No. pits	Pit size (μm^2)	*M* (%)
1	$1.8 \cdot 10^5$	173.6	4.124	3.67
	$3.6 \cdot 10^5$	314.04	9.708	15.90
2	$1.8 \cdot 10^5$	96.8	15.484	7.41
	$3.6 \cdot 10^5$	112	20.88	12.24
	$5.4 \cdot 10^5$	126.4	32.08	21.34
3	$1.8 \cdot 10^5$	34.4	2.25	0.42
	$3.6 \cdot 10^5$	30	2.42	0.39
	$5.4 \cdot 10^5$	101.75	3.59	2.10
	$7.2 \cdot 10^5$	38.4	21.132	3.67
4	$1.8 \cdot 10^5$	176.6	6.728	6.56
	$3.6 \cdot 10^5$	145	21.02	13.34
	$5.4 \cdot 10^5$	86.5	36.59	17.59
5	$1.8 \cdot 10^5$	112.5	5.90	3.69
	$3.6 \cdot 10^5$	141.6	14	9.46
6	$1.8 \cdot 10^5$	119.5	28.65	15.89
7	$1.8 \cdot 10^5$	27.4	38.242	5.26
	$3.6 \cdot 10^5$	27.6	76.66	10.34
8	$1.8 \cdot 10^5$	144.75	6.128	4.97
	$3.6 \cdot 10^5$	135.2	9.422	6.93
	$5.4 \cdot 10^5$	170.25	11.62	10.82
	$7.2 \cdot 10^5$	120	27.53	14.09

It should be noted that when this method is applied to an untested surface the micropitting parameter is not zero. M (%) has been determined for the discs before test and the average value is $M = 0.5 \pm 0.3$ %. This is, on one hand because small pits randomly distributed, situated on the top of the grinding marks pre-exist due to the surface finishing process and, on the other hand, because the limitations of image processing method. The aforementioned considerations suggest that there is a threshold value for M below which the method is not sensitive. The threshold value was chosen as $M = 1.5$ %. The results of the micropitting measurements are given in Table 5-8. These data were used to plot the micropitting parameter versus the number of stress cycles in Figure 5-13. As explained in section 4-2, the two-disc machine cannot be run with very rough specimens. In experiment 6, after the first testing stage, the micropitting has developed to such an extent that the test needed to be stopped. In order to be able to plot a curve, the value of micropitting parameter $M = 0.5$ %, measured before test has been used. Because, in experiment 3, the first two measurements give values of the micropitting parameter smaller than the threshold (< 1.5 %), these two values were omitted.

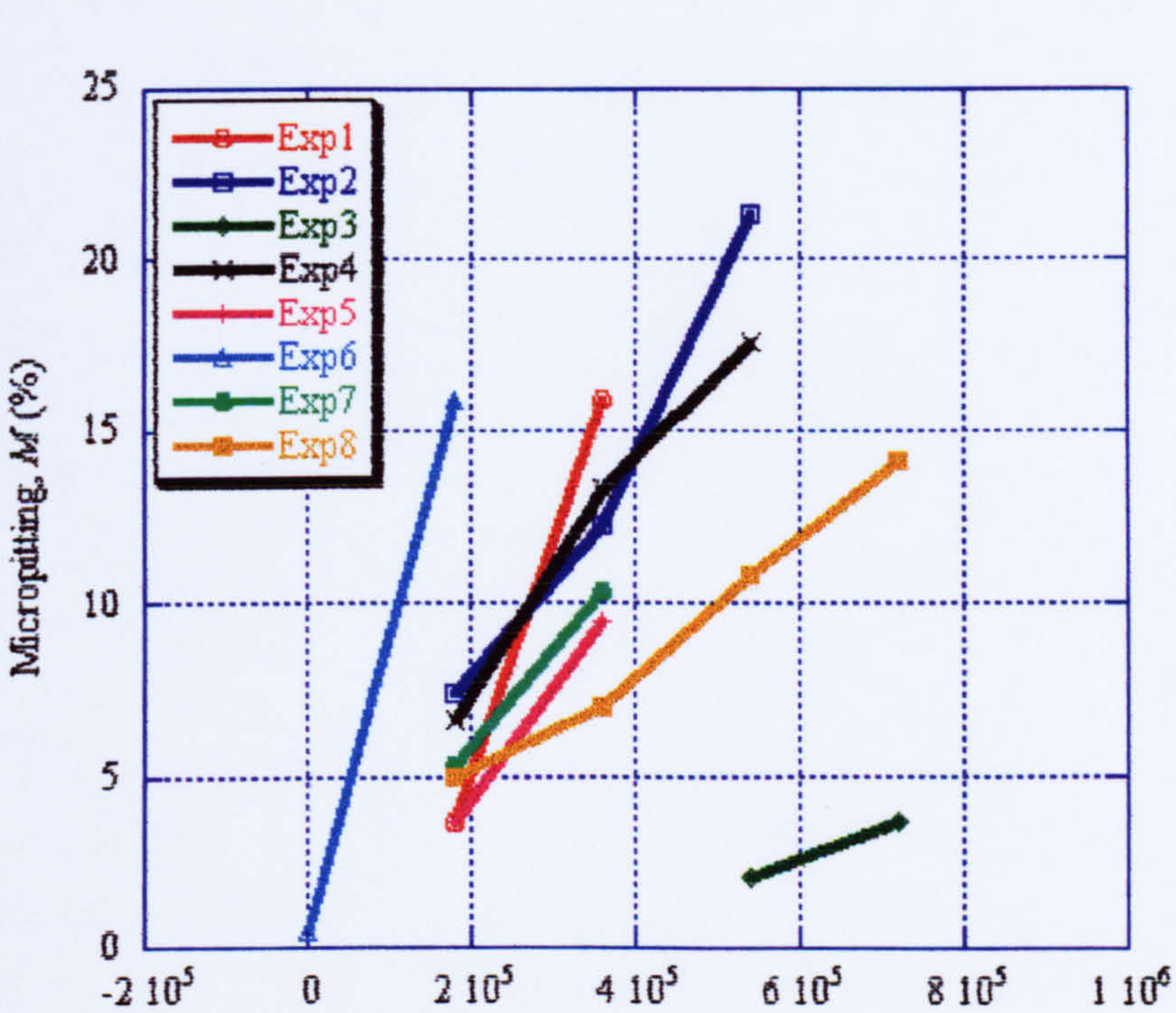


Figure 5-13. Percentage micropitting versus number of cycles.

The curves from Figure 5-13 have been fitted to lines in Figure 5-14. These curves show the rate of micropitting propagation and they follow the same pattern as the wear rates from Figure 5-9. This suggests that the removal of material from the specimen surface is mainly due to micropitting. It should be noted that the curves from Figure 5-14 are better separated one from another than the curves from Figure 5-9. Based on this, it can be assumed that the micropitting measurements have better accuracy than the measurement of mass losses.

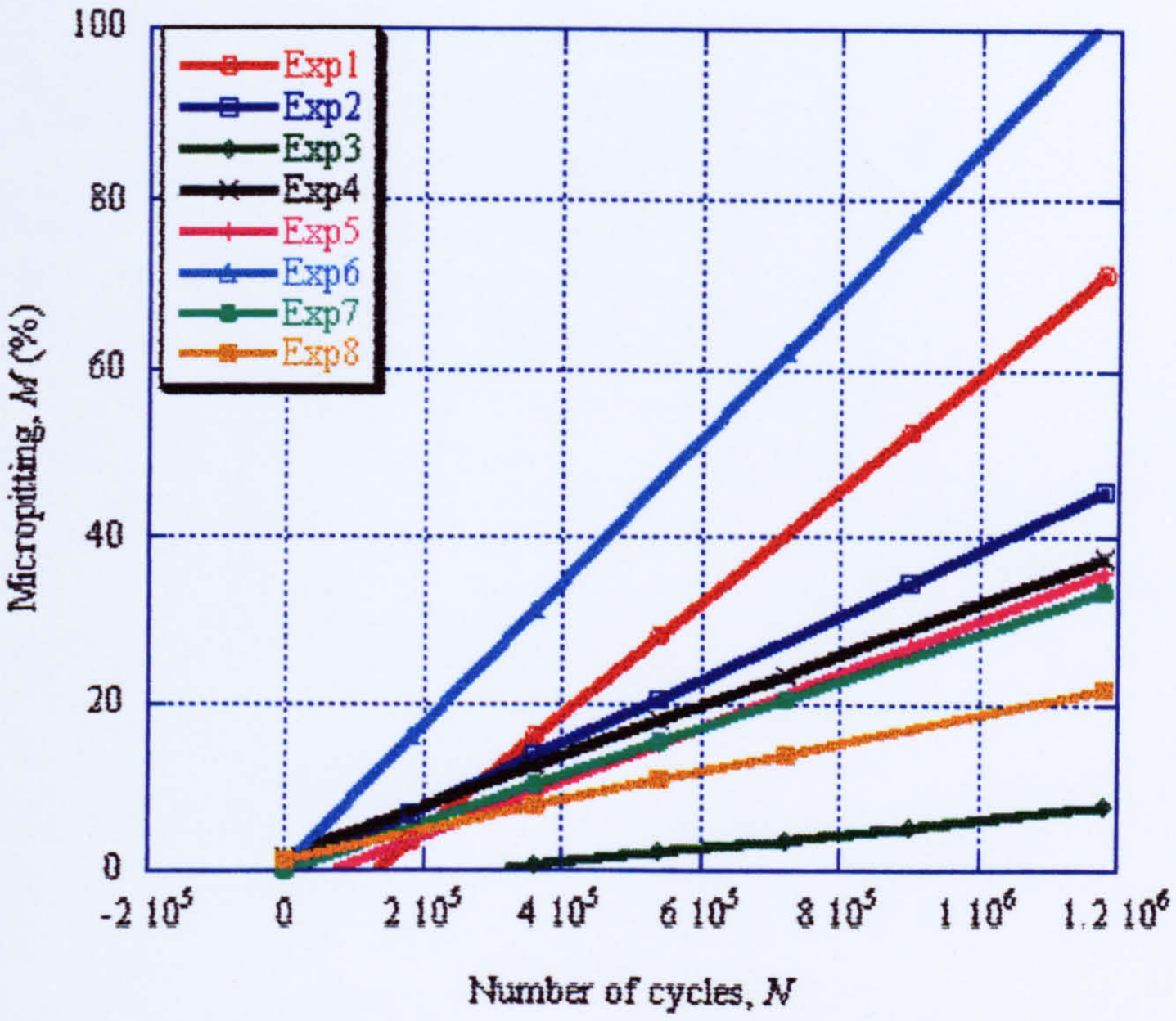


Figure 5-14. Line fitting of the micropitting curves.

It is now possible to investigate the micropitting initiation and propagation. The number of cycles, N_0, necessary for micropitting initiation is taken to be the interception of the micropitting curve with a horizontal line. As already discussed, this line is represented by the threshold $M = 1.5$ %. The slopes, dM/dN, of the lines from Figure 5-14 represent the *micropitting propagation rates*. The number of cycles for micropitting initiation, N_0, and the micropitting propagation rates, dM/dN, determined for the eight experiments are given in Table 5-9. N_0 has been plotted as columns in Figure 5-15 and dM/dN in Figure 5-16.

It can be seen that the initiation and propagation of micropitting do not follow a similar pattern. As a function of the number of cycles, N_0, necessary for micropitting to initiate,

the eight experiments can be classified as follows: $N_0 \approx 10^3$ cycles – experiment 4 and 8, $N_0 \approx 10^4$ cycles – experiments 2, 6 and 7, and $N_0 \approx 10^5$ cycles – experiments 1, 3 and 5. The analysis of these results will be presented in Chapter 8.

Table 5-9. Micropitting initiation and propagation.

Experiment	Initiation cycles, N_0 (M=1.5%)	Propagation rate, dM/dN
1	$1.48 \cdot 10^5$	$6.80 \cdot 10^{-5}$
2	$4.55 \cdot 10^4$	$3.87 \cdot 10^{-5}$
3	$4.72 \cdot 10^5$	$8.77 \cdot 10^{-6}$
4	$1.05 \cdot 10^3$	$3.06 \cdot 10^{-5}$
5	$1.12 \cdot 10^5$	$3.21 \cdot 10^{-5}$
6	$1.16 \cdot 10^4$	$8.55 \cdot 10^{-5}$
7	$4.69 \cdot 10^4$	$2.83 \cdot 10^{-5}$
8	$6.24 \cdot 10^3$	$1.74 \cdot 10^{-5}$

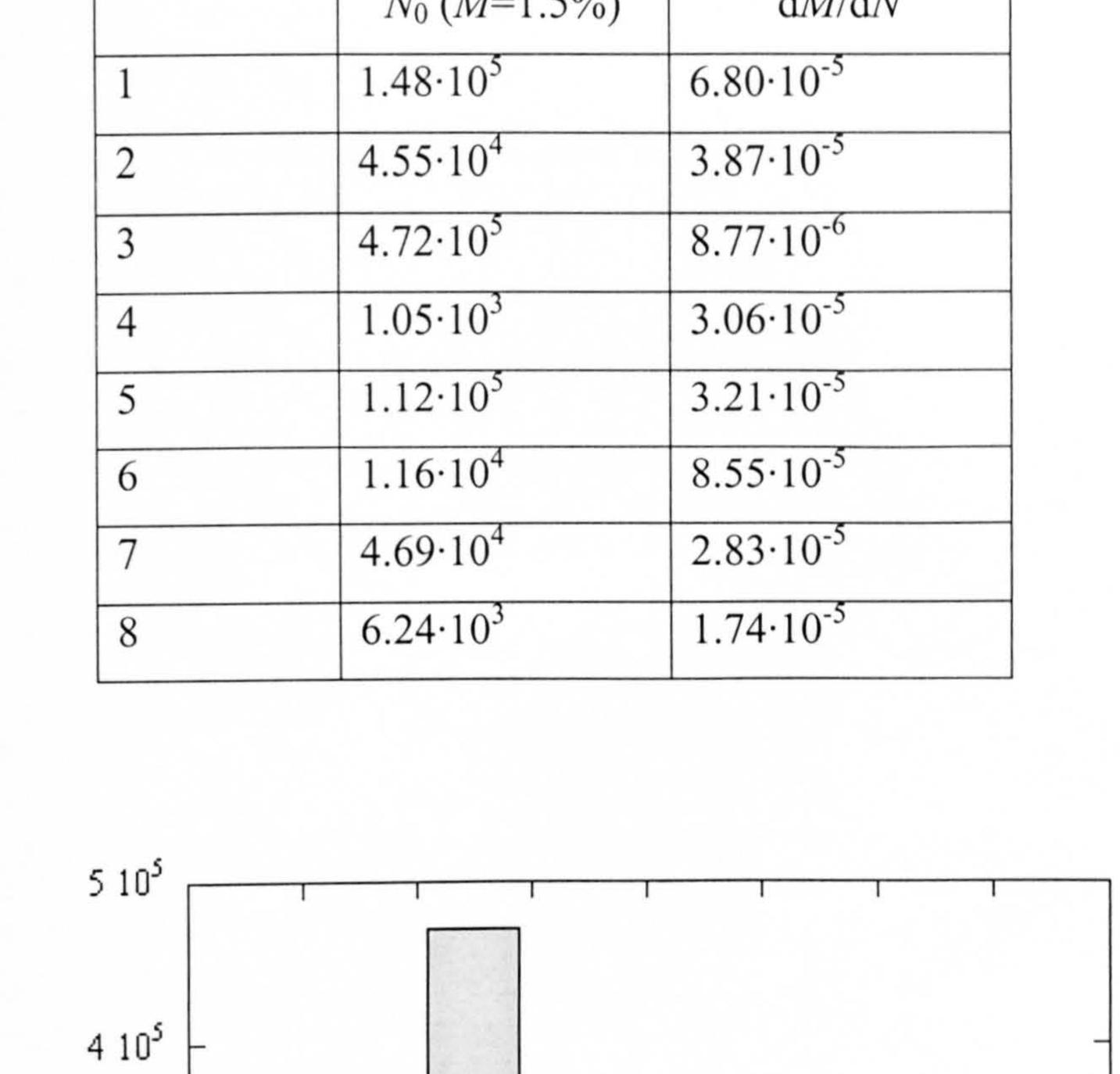


Figure 5-15. Micropitting initiation.

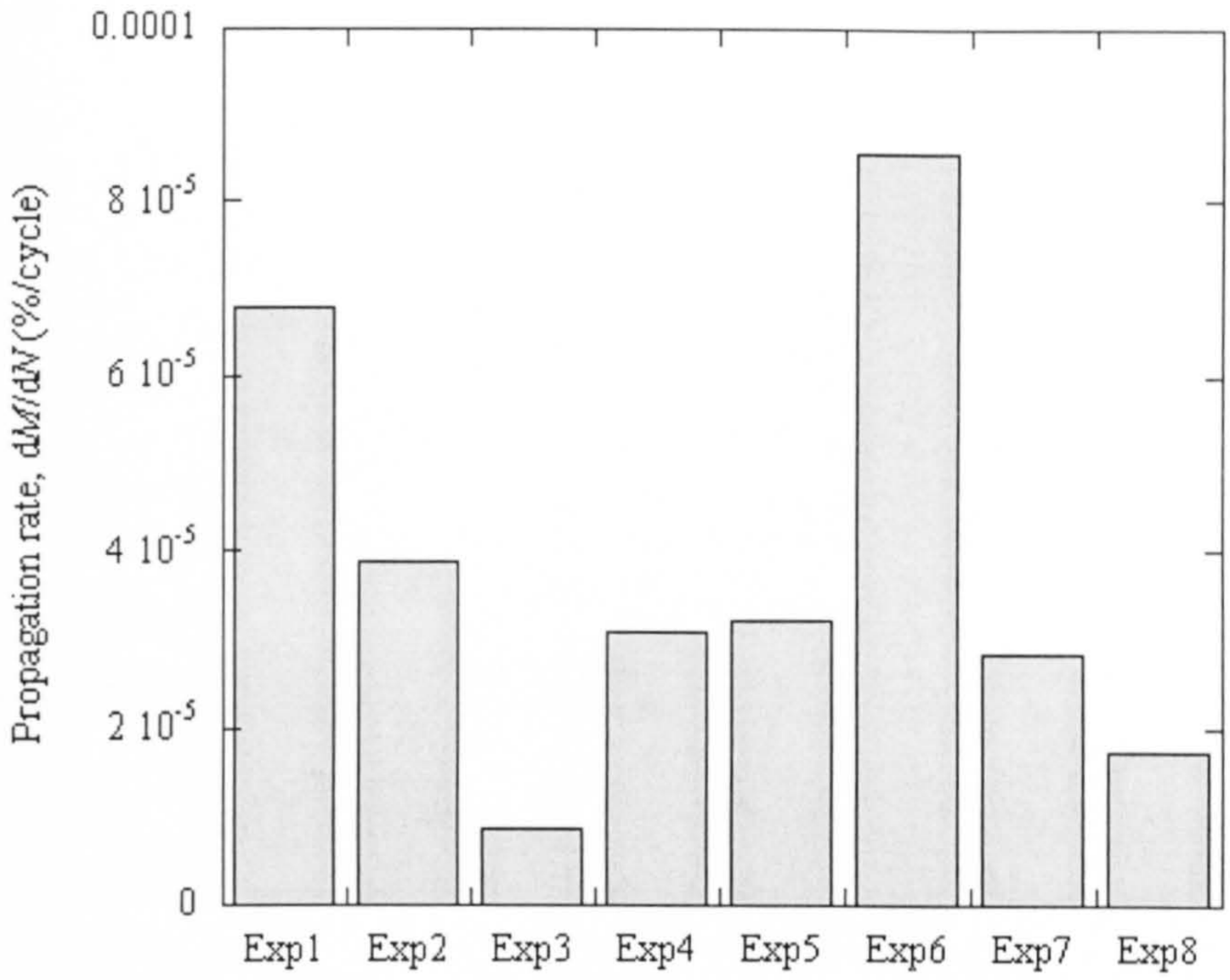


Figure 5-16. Micropitting propagation.

5.7 Initiation and propagation of cracks

As any other fatigue phenomena, micropitting is a process driven by crack initiation and propagation. From the point of view of the location of the initiation site, the cracks observed in this study can be classified in two groups, as follows: cracks initiated below the surface also known as *subsurface cracks* (see Figure 5-17a) and cracks initiated at the surface or *surface cracks* (see Figure 5-17b). In what follows these two categories are treated separately.

5.7.1 Subsurface cracks

The cracks initiated in the subsurface regions tend to propagate in a plane parallel to the surface plane for significantly long distances (see Figures 5-17a, 5-18 and 5-19). Considering the size of the micropits (i.e., in the order of microns) and the length of the observed subsurface cracks (tens to hundreds of microns) it can be concluded that these cracks are not responsible for micropitting initiation and propagation but they probably lead to a delamination wear process. However, some of these subsurface cracks can

propagate towards the surface, as in Figure 5-20a or they can intersect with a crack originated at the surface as in Figure 5-20b. In both cases they can cause a micropit, although the features from Figure 5-20 are rather an exception in this study.

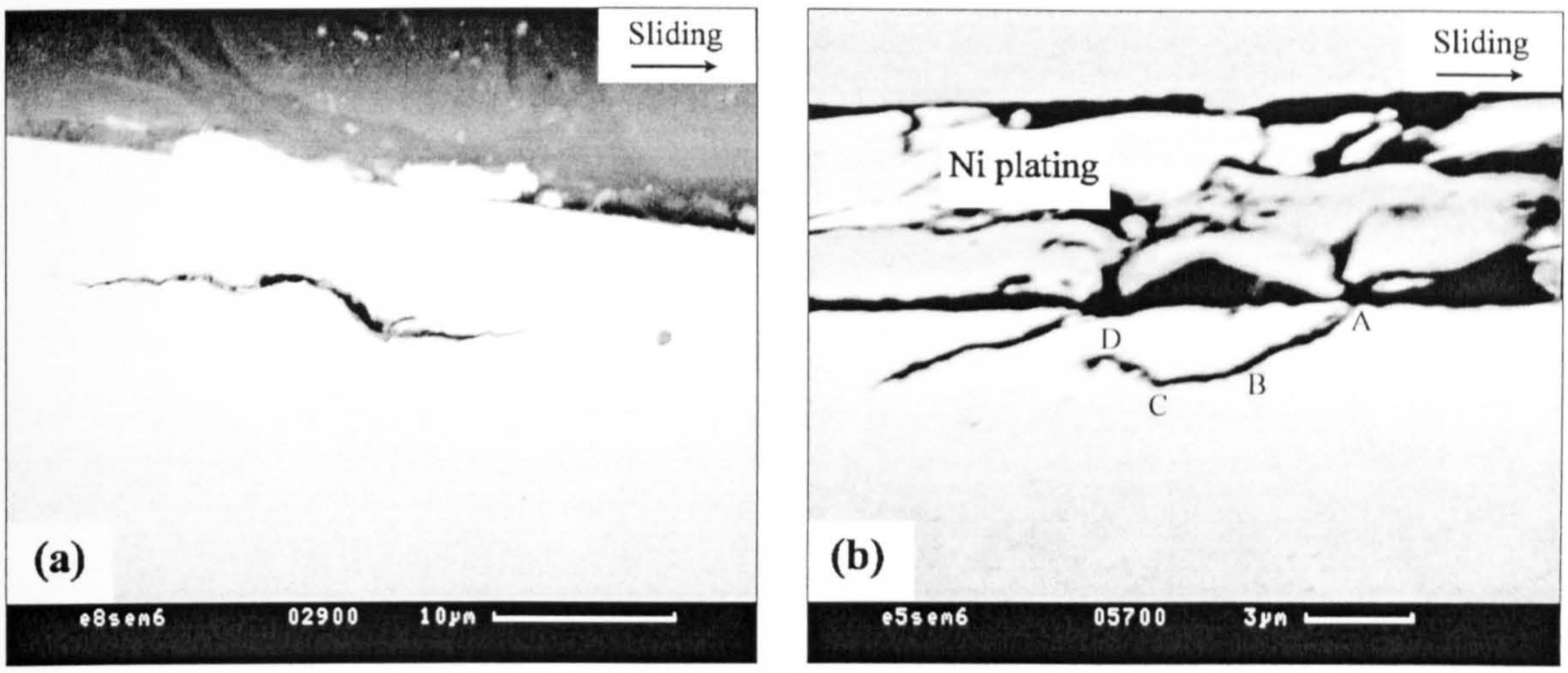


Figure 5-17. Cracks initiated (a) below the surface and (b) at the surface. Backscattered electron images taken on longitudinal cross sections from specimen 8 (a) and 5 (b). Surface lay longitudinal.

The cracks shown in Figure 5-17a and 5-18 follow a direction that is parallel to the surface lay and the depth below the surface is in the range of the depth of maximum shear stress induced by asperities, z_i. These considerations indicate that a crack occurs below the grinding marks due to asperity contact and propagates due to contact between the neighbouring asperities.

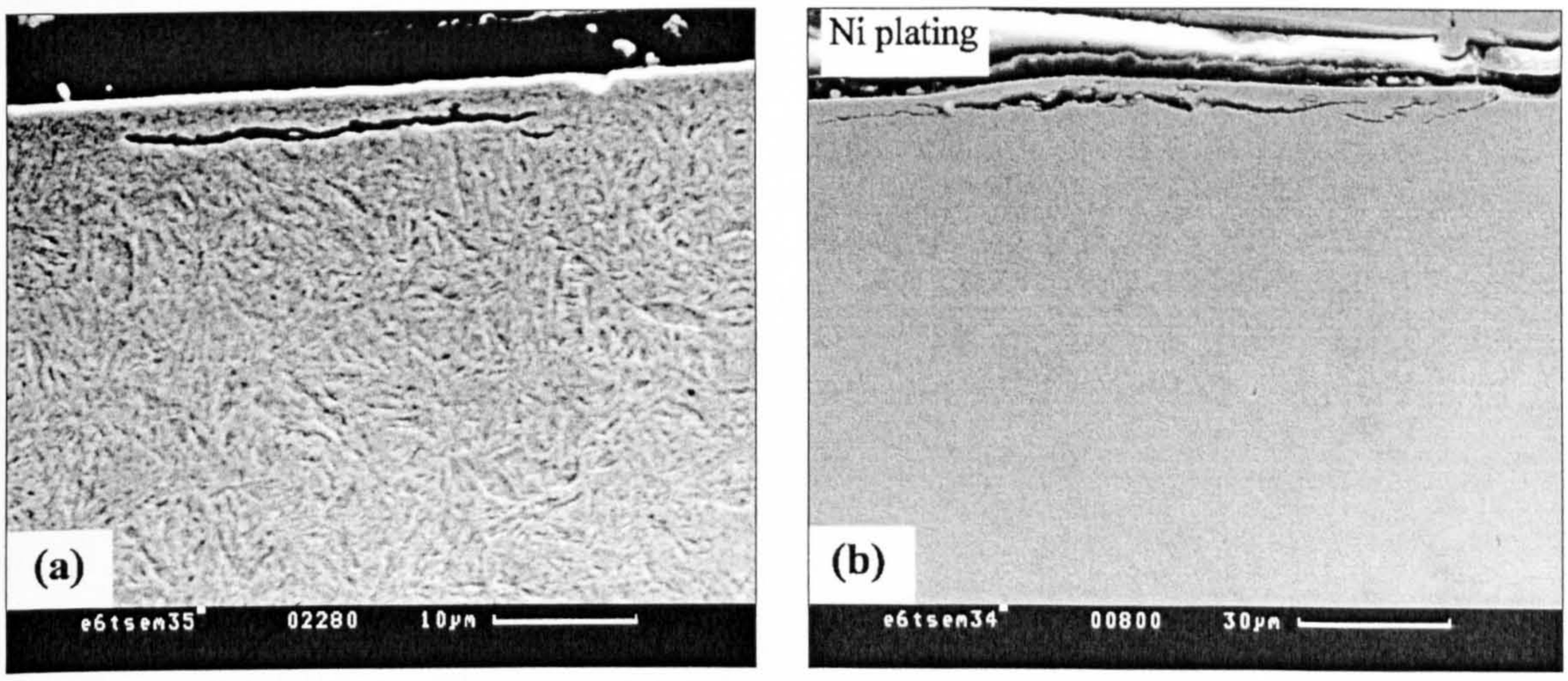


Figure 5-18. Cracks parallel to the surface. Secondary electron images taken on transverse cross sections from specimen 6. Etch: Nital 2%. Surface lay transverse.

The depth of the parallel cracks varies, from less than one micron (Figure 5-19a) to about 20 microns (Figure 5-19b). Long, parallel cracks propagating very close to the surface are commonly observed in all specimens. This suggests that the subsurface region up to a depth of 5-6 μm represents a critical region for the initiation and propagation of microcracks.

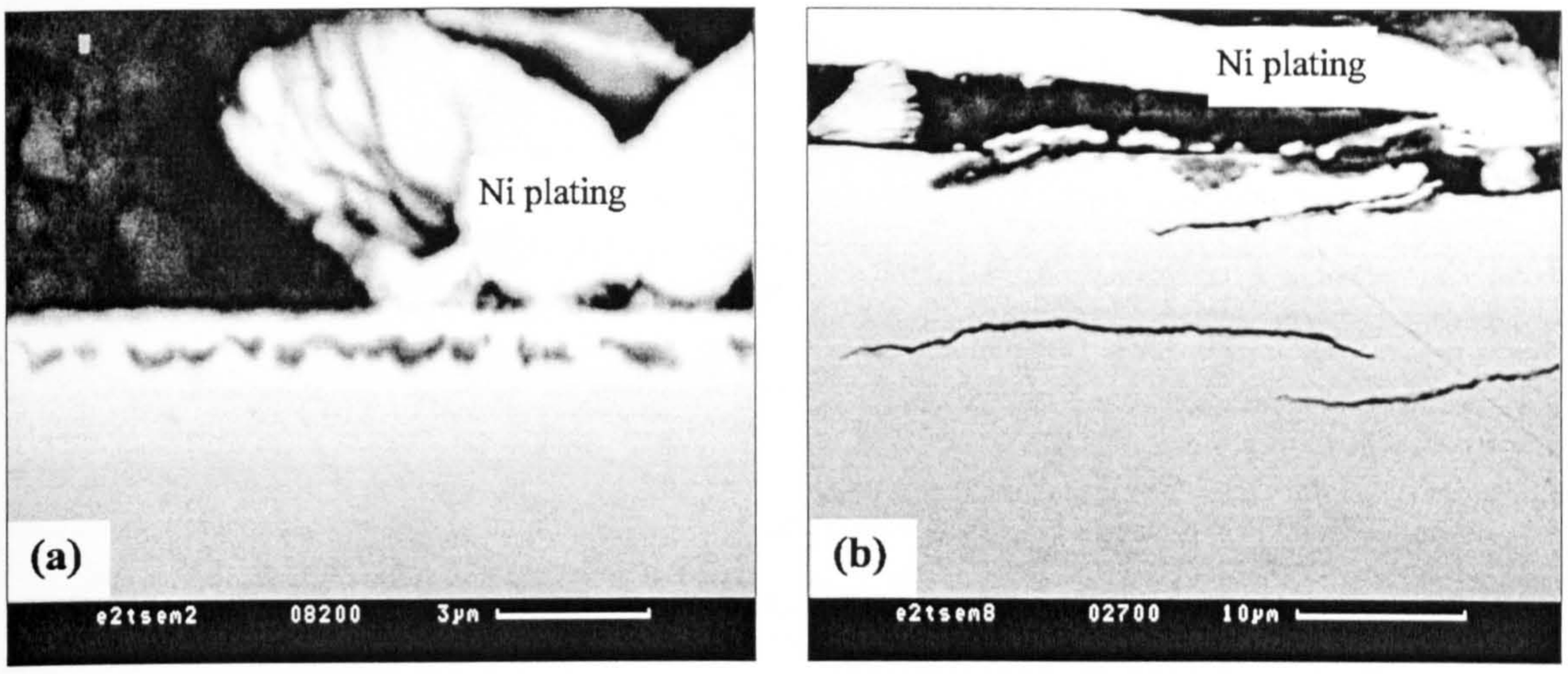


Figure 5-19. Cracks parallel to the surface. Backscattered electron images taken on transverse cross sections from specimen 2. Transverse lay.

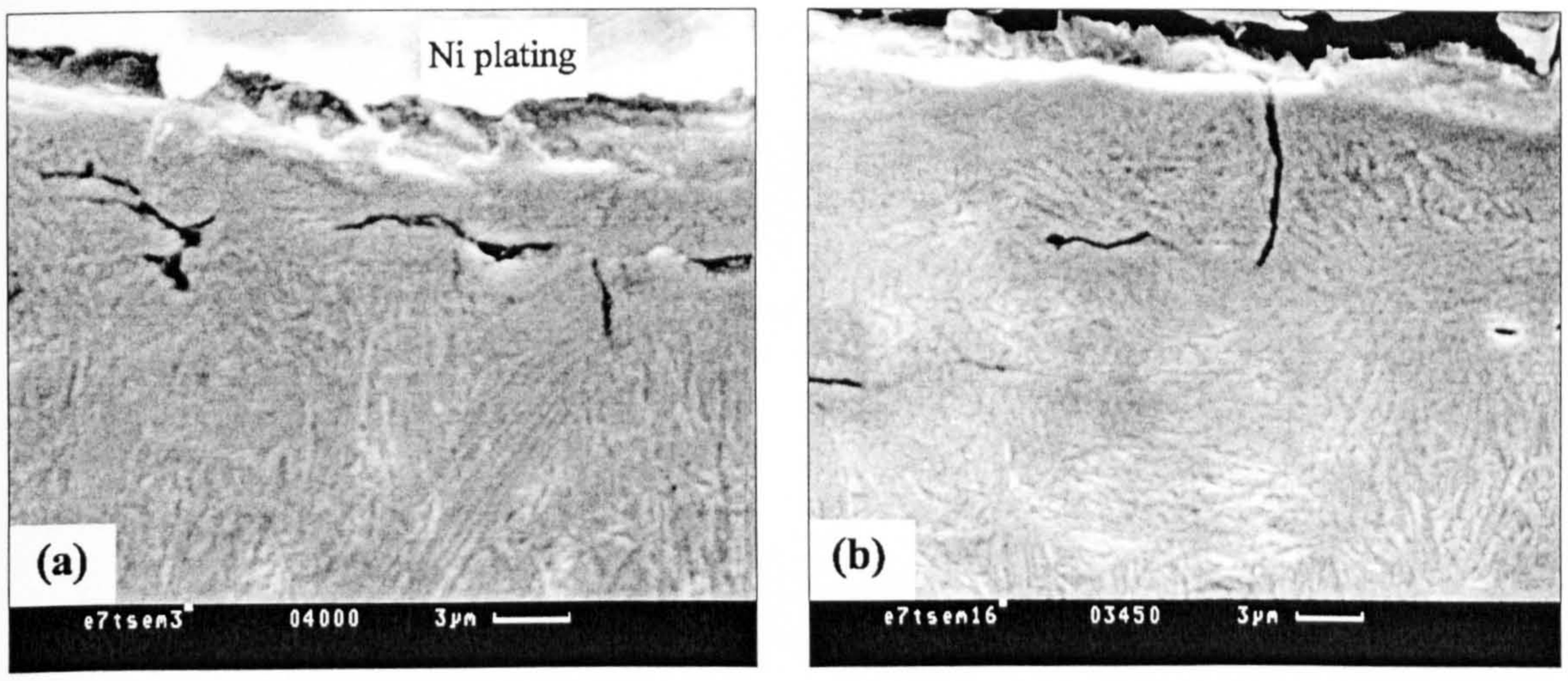


Figure 5-20. Subsurface cracks. Secondary electron images taken on transverse cross sections from specimen 7. Etch: Nital 2%. Longitudinal lay.

5.7.2 Surface cracks

The cracks initiated at the surface propagate in a plane which is inclined to the surface plane at a shallow angle (Figure 5-17b). After a growth of few microns (AB) the crack

turns parallel to the surface plane (BC) and after some growth changes direction towards the surface (CD). When the crack reaches the surface a micropit is formed (marked A in Figure 5-21a & b). In this case the initial crack continues to propagate into the bulk and new branches oriented towards the surface form at different depths (B and C in Figure 5-21b). The formation of multiple branches suggests the following mechanism of micropitting propagation.

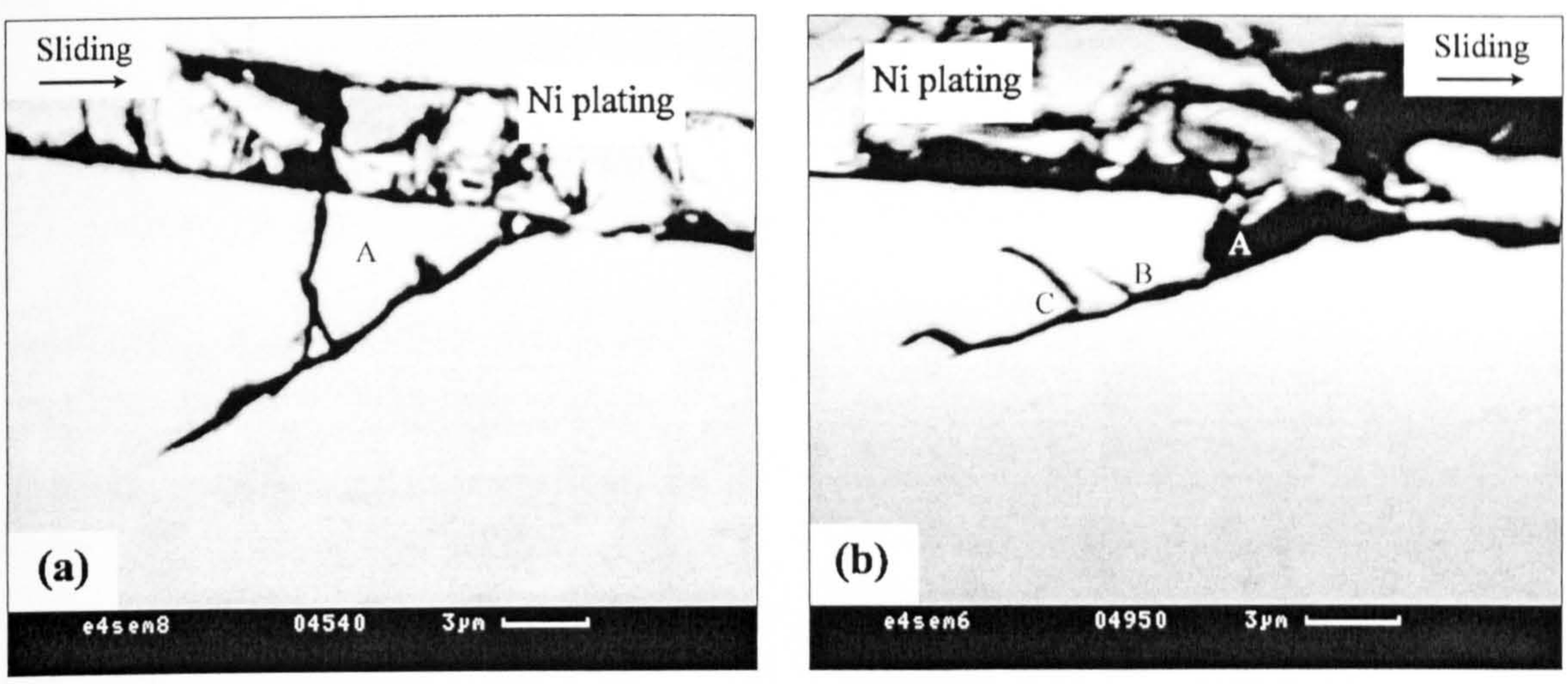


Figure 5-21. Crack propagation leading to the formation of micropits. Backscattered electron images taken on longitudinal cross sections from specimen 4 (surface lay transverse).

Figure 5-22a shows a crack initiated at the surface, in A, it propagates at a shallow angle and in B branches occur, one branch (BC) propagates towards the surface to form a micropit and the other branch continues to propagate into the bulk.

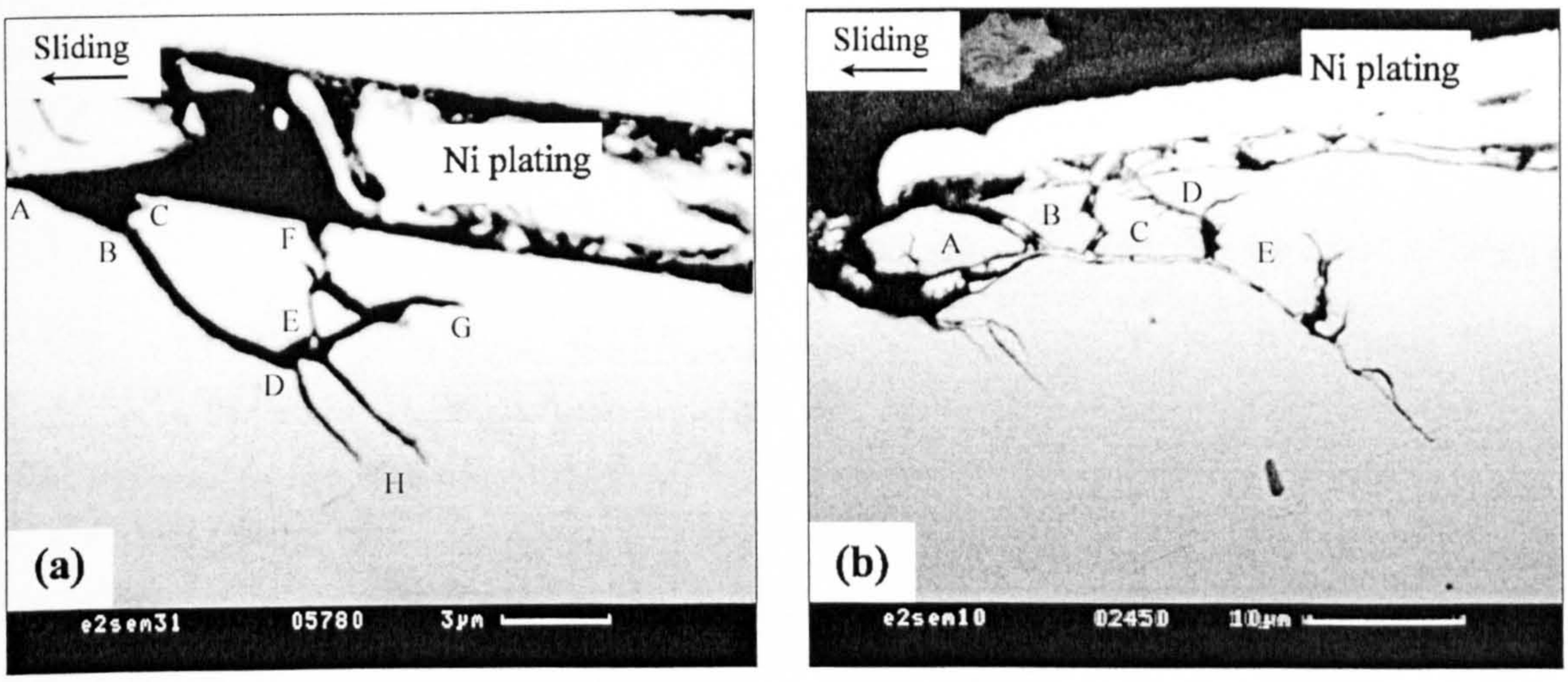


Figure 5-22. Crack branching and micropitting propagation. Backscattered electron images taken on longitudinal cross sections from specimen 2 (surface lay transverse).

After some growth (BD), the branching occurs again, towards the surface (DEF and DEG) and into the bulk (DH). In this way the depth and the volume of the micropits increases. In Figure 5-22b the initial pit A increases its volume by adding new volumes of material removed from the surface B, C, D and E. The crack branching follows a complex pattern, as shown in Figure 5-23, giving rise to a complicated network of cracks. It is suggested that the crack propagation and branching depends on material microstructure. The intersection of cracks will lead to material weakening in the regions below the surface.

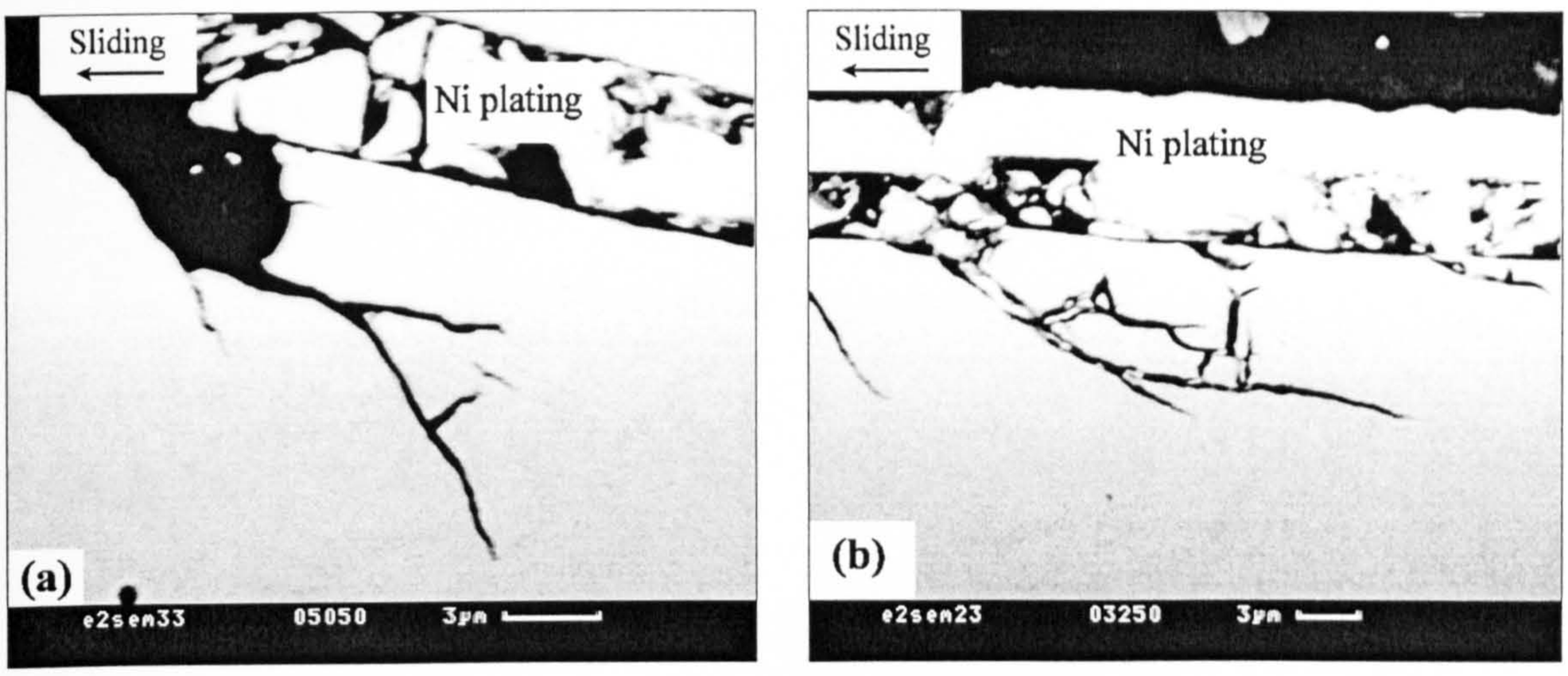


Figure 5-23. (a) Intersection of cracks, and (b) network of cracks below the contact surface. Backscattered electron images taken on longitudinal cross sections from specimen 2.

Table 5-10. The predicted depth of maximum shear stress due to asperity contact, z_i, the average depth of observed cracks, d, and their angle of propagation, α.

Experiment	Depth of maximum shear stress, z_i (μm)	Depth of cracks, d (μm)	Crack angle, α (°)
1	7.16 -14.51	11.26	30.30
2	11.73 - 15.19	12.22	31.31
3	4.33 - 11.66	6.16	60.30
4	6.21 - 6.71	6.32	28.05
5	3.90 - 8.56	4.79	19.96
6	5.04 - 9.08	7.67	42.72
7	5.28 - 7.44	7.02	24.26
8	9.30 - 16.74	7.72	16.10

Table 5-10 summarises the results of the measurements of the average depth of cracks, d, and the average angle of propagation, α. The given values represent the average of the measurements done on ten cracks for each specimen. The measured depth is compared with the range of depth of the maximum shear stress, z_i. It can be seen that the depth of cracks, d, lies between the minimum and maximum depth z_i. It follows that the cracks initiate at the surface and propagate in the subsurface regions where the shear stress is maximum.

5.8 Residual stresses and retained austenite

Residual stresses have been measured at the surface in two orthogonal directions. As shown in Figure 5-24 the compressive residual stresses increase after the first stage of running and they remain almost constant during the following stress cycles. The first measurements have been taken after $1.8 \cdot 10^5$ cycles, but changes in residual stress probably occur immediately after starting the test due to plastic deformation. Since the variation of residual stress is not significant, they seem to have little effect on micropitting.

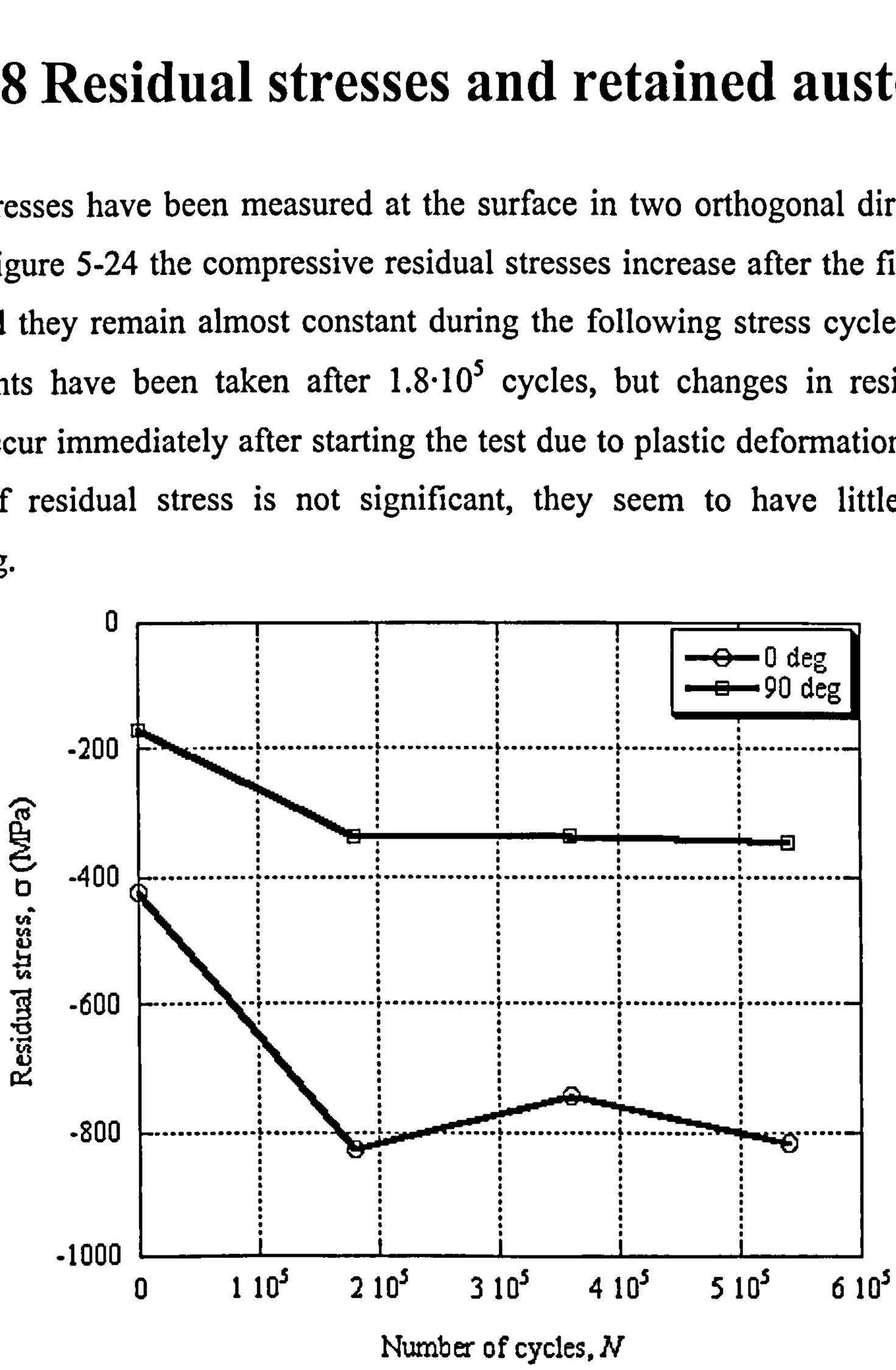


Figure 5-24. Variation of residual stress during testing. Experiment 8.

Similarly, the amount of retained austenite drops from 15 % to a value of 7 %, which remains effectively constant for most of the tests (see Figure 5-25).

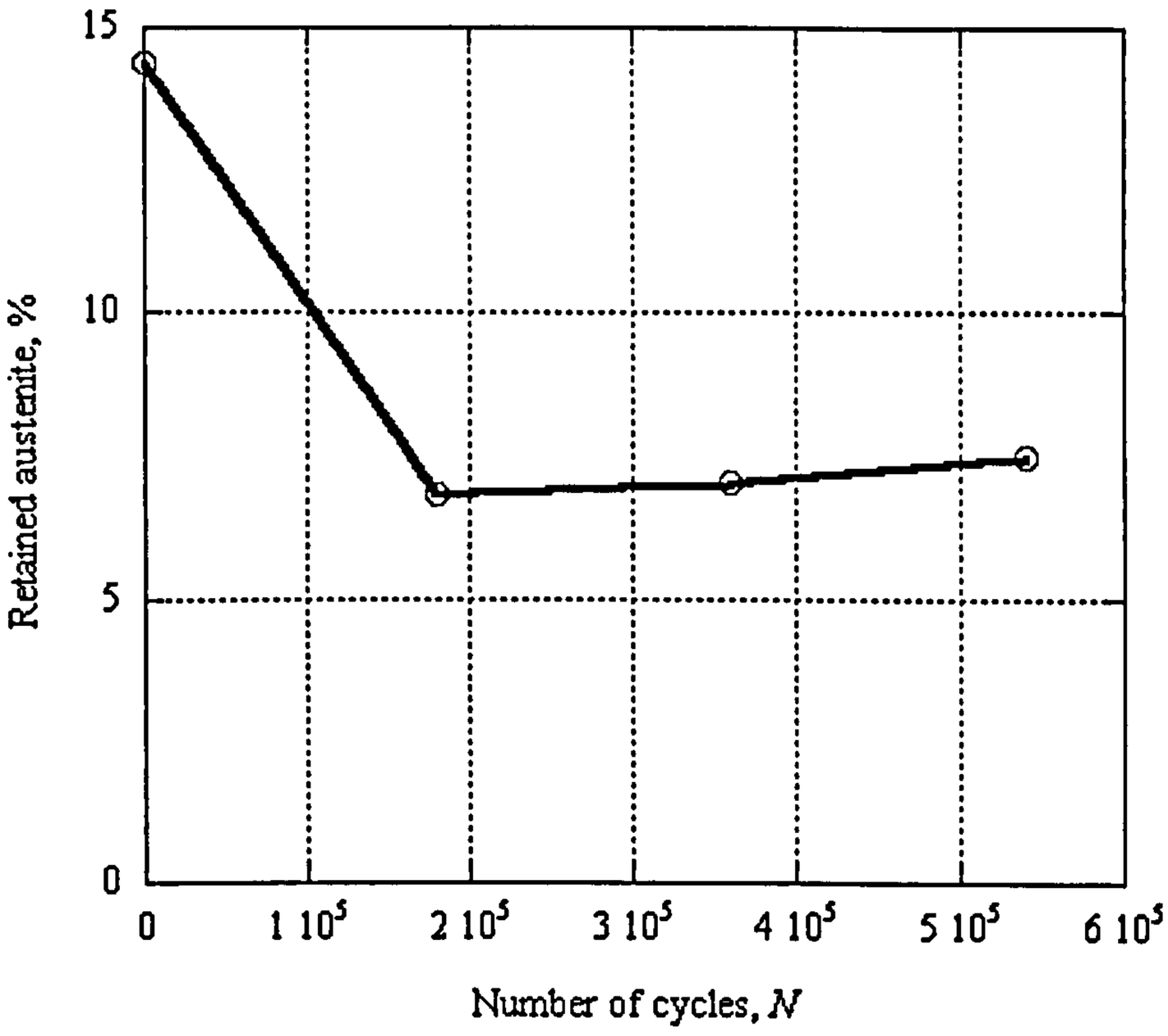


Figure 5-25. Variation of the retained austenite percentage. Experiment 8.

5.9 Hardness of the testing surface

It has been shown so far that the plastic deformation of asperities plays a key role in the development of micropitting. Therefore, the hardness of the steel at the surface and how this changes during running is a very important factor but the surface roughness makes hardness measurements difficult. It is possible to find flat areas on the specimen surface but they are very small in size, usually less than ten microns. It follows that microhardness testing with relatively low loads can be used to determine the hardness of the test surface. In this work, the Vickers microhardness tests were carried out with 25g load. The resulting indentations have diagonals of approximately 7-8 microns. The diagonals have been measured using optical profilometer so the effect of pile up has been eliminated. Figure 5-26 shows an indentation performed on the rough surface of the disc tested in experiment 1. As previously discussed, the disc surface tends to flatten

during the test so the inconvenience posed by surface roughness in hardness measurements is more pronounced when the test is done on specimens before running. The results of microhardness measurements are given in Table 5-11 for four tests. In Figure 5-27 the surface hardness has been plotted against the number of cycles.

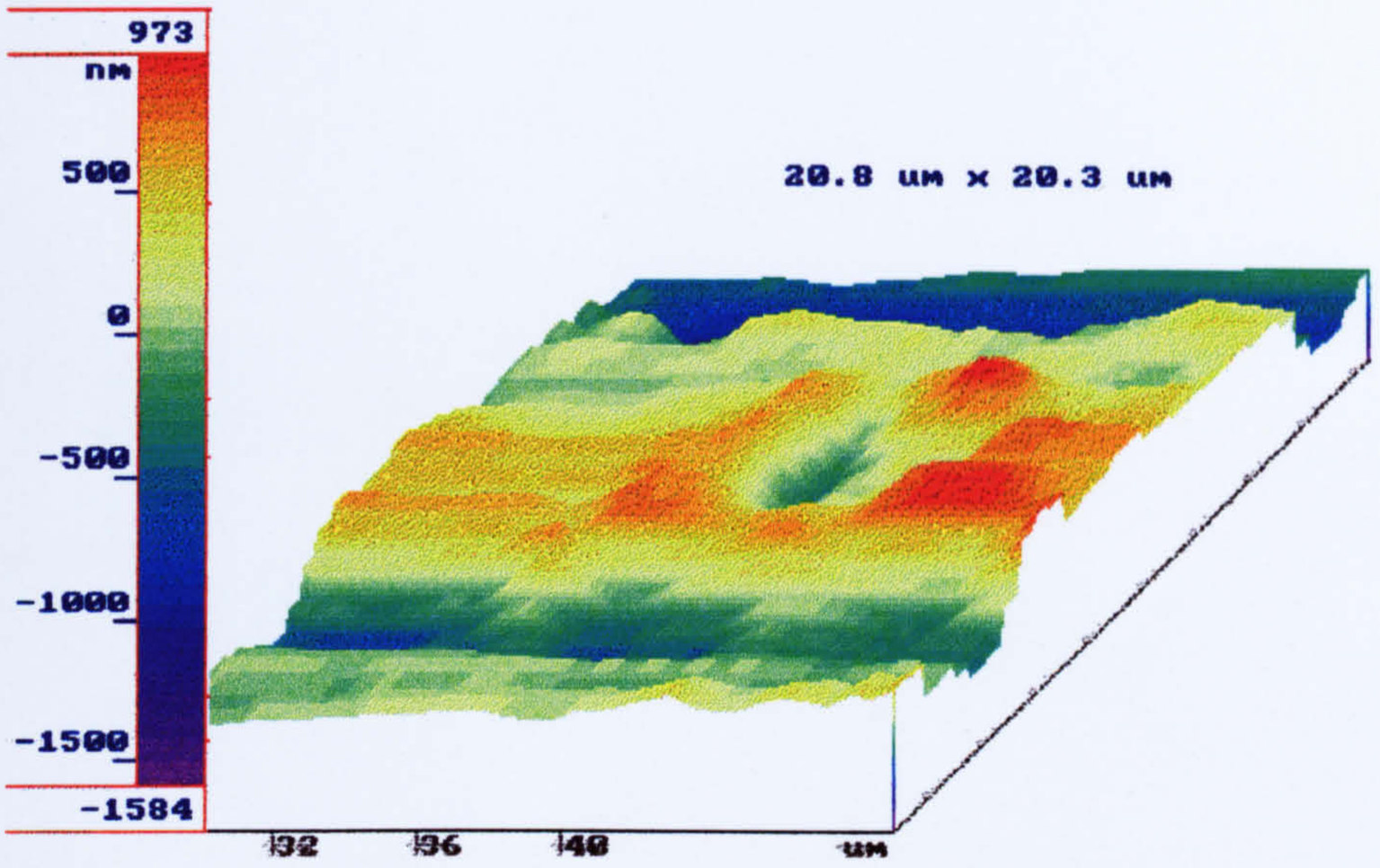


Figure 5-26. Optical profilometer image of a 25g indent.

The plots from Figure 5-27 show an increase in hardness during the period of micropitting initiation and a drop after micropitting has developed to some extent. The increase in hardness can be related to the transformation of retained austenite and to a work hardening phenomenon due to plastic deformation. The decrease in hardness is probably linked to microstructural changes. Due to scatter in the measured data it is not possible to confirm these hypothesis by hardness testing alone.

Table 5-11. Surface micro-hardness.

Number of cycles, N	Hardness, H_v (GPa)			
	Exp1	Exp3	Exp6	Exp7
0	9.23	8.80	9.10	8.85
$1.8 \cdot 10^5$	10.3	8.27	8.89	9.58
$3.6 \cdot 10^5$	9.28	9.54	-	8.11
$5.4 \cdot 10^5$	-	7.72	-	-
$7.2 \cdot 10^5$	-	7.86	-	-

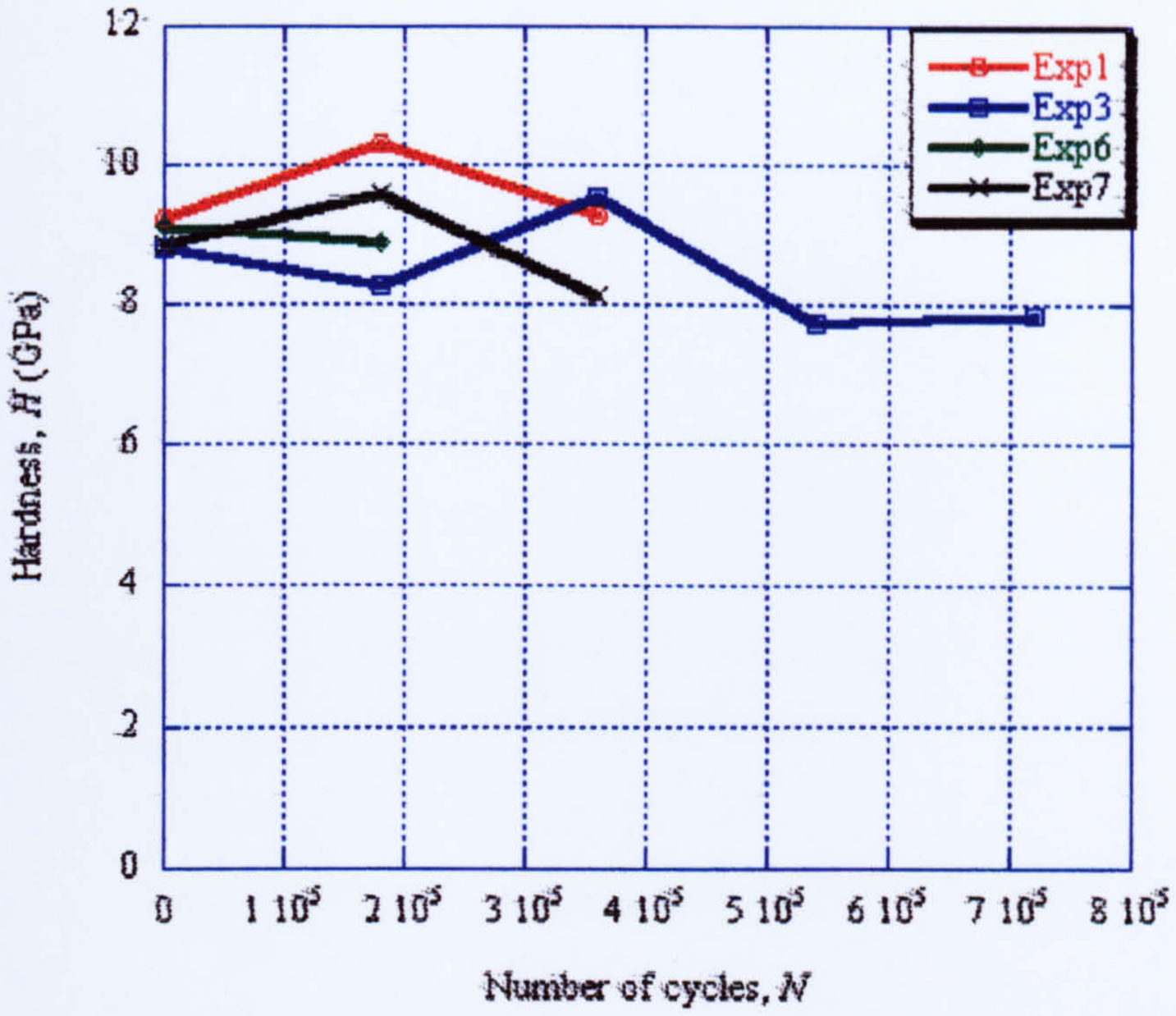


Figure 5-27. Surface hardness vs. number of stress cycles.

Even at very low loads, the Vickers indenter produces relatively large indentations such that the hardness determined represents the composite hardness of the different phases that are present in the steel (i.e., martensite and retained austenite). The nanoindentation technique offers the possibility to determine the hardness of individual phases. The surface of the disc specimen has been tested by nanoindentation after the micropitting experiments were stopped. The indents were performed in areas flattened during cycling so the effect of surface roughness is diminished. The hardness lies in relatively large interval of values. It is more convenient to analyse the data in a statistical way rather than presenting the mean value. The cumulative distribution of the hardness values obtained for the discs tested in experiment 1 and 6 have been plotted in Figure 5-28 and 5-29, respectively.

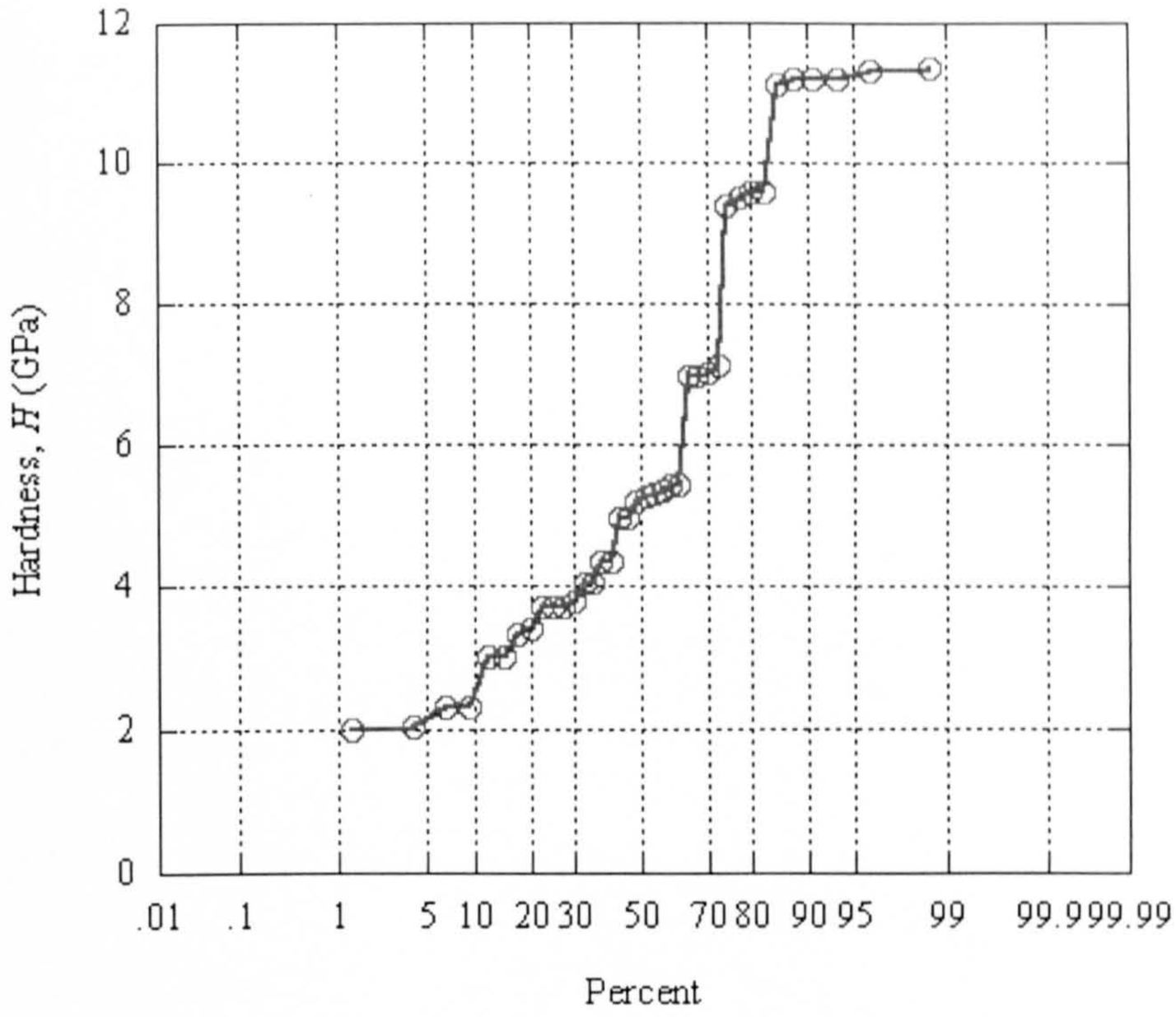


Figure 5-28. Cumulative probability plot of the nanohardness data. Experiment 1.

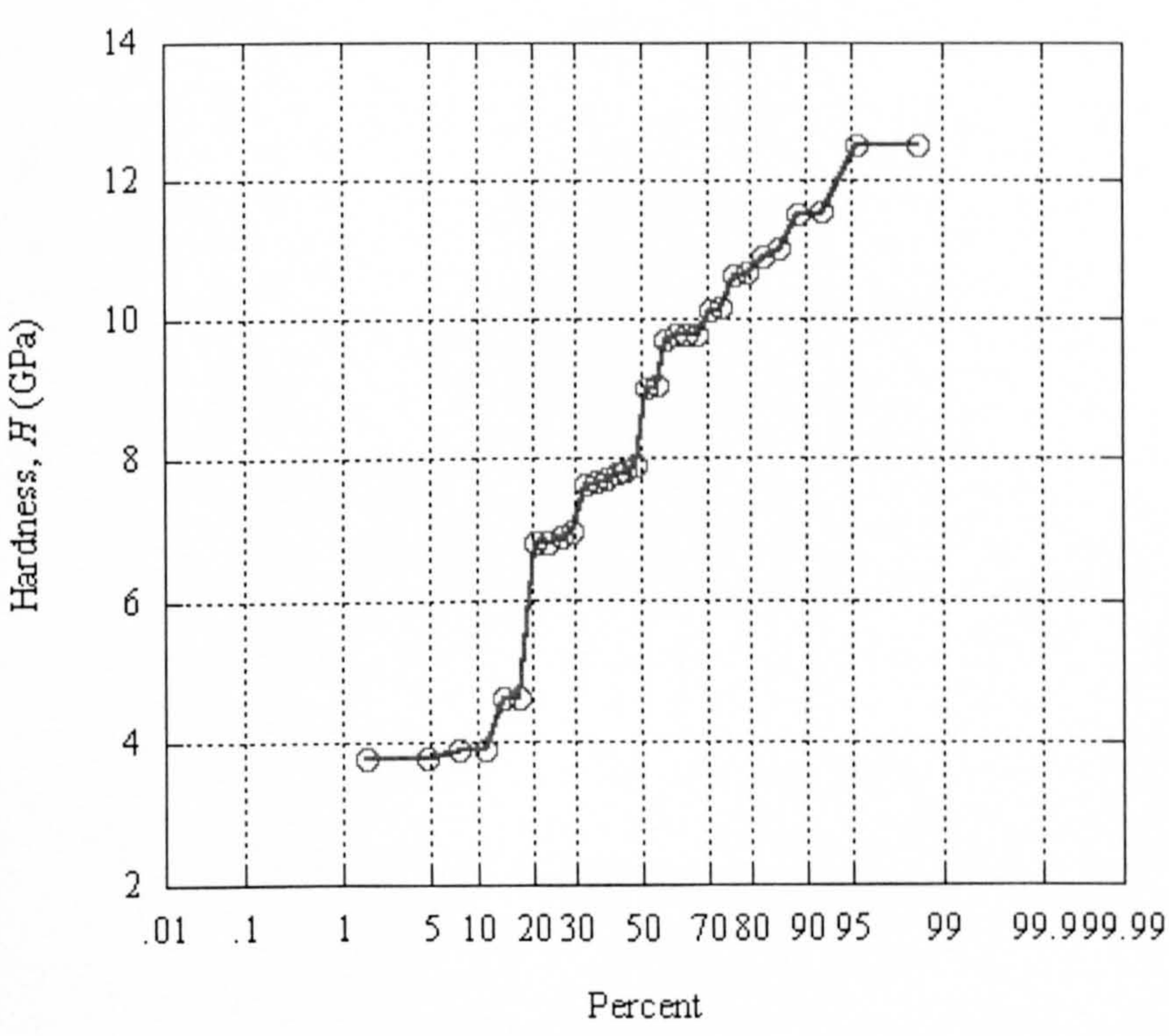


Figure 5-29. Cumulative probability plot of the nanohardness data. Experiment 6.

Both graphs depict two distinct distributions of the hardness values, corresponding to two phases, a soft phase with hardness below 6 GPa and a harder phase above 6 GPa.

The soft phase is represented by the retained austenite and the hard phase by martensite. The average values of the data below 6 GPa are $H_{RA} = 3.95 \pm 1.13$ GPa in Figure 5-28 and $H_{RA} = 4.10 \pm 0.41$ GPa in Figure 5-29 and they represent the hardness of the retained austenite. The average values of the data above 6 GPa are $H_M = 9.52 \pm 1.81$ GPa in Figure 5-28 and $H_M = 9.32 \pm 1.78$ GPa in Figure 5-29 and they represent the hardness of the martensite. Considering the amount of retained austenite determined by X-ray diffraction, which according to Figure 5-25 is 7 % the composite hardness can be calculated with the expression given below

$$H = f_{RA} \cdot H_{RA} + f_M \cdot H_M \tag{5-5}$$

where f_{RA} and f_M are the volume fraction of retained austenite and martensite.

If the terms in (5-5) are replaced by the average hardness values calculated above, then

$$H(\text{Exp1}) = 0.07 \cdot 3.95 + 0.93 \cdot 9.52 = 9.13\,\text{GPa} \tag{5-6}$$

$$H(\text{Exp6}) = 0.07 \cdot 4.10 + 0.93 \cdot 9.32 = 8.95\,\text{GPa} \tag{5-7}$$

The hardness obtained with relations 5-6 & 7 corresponds well to the hardness determined with Vickers tester (see Figure 5-27).

5.10 Summary

The examination of the surfaces affected by micropitting has revealed that the formation of micropits follows a spatial distribution, which can be approximated by the surface profile wavelength in two orthogonal directions: parallel and perpendicular to the surface lay. Based on this observation it is suggested that micropitting occurs due to asperity contact. The wavelength ratio, Λ (the ratio of the surface roughness wavelength in the direction perpendicular to the direction of the contact motion, λ_{qy}, to the surface roughness wavelength in the direction parallel to the direction of the contact motion, λ_{qx}) has been assigned as the most suitable parameter for the characterisation of the surface roughness.

The surface inside micropits shows the characteristics of a surface resulting from ductile fracture, which has undergone plastic deformation prior to fracture.

The Greenwood-Williamson model has been used to determine the asperity parameters. The depth of maximum shear stress induced by asperity lies in the range 3 to 15 μm below the surface, which corresponds to the depth of crack propagation. As a result, it is suggested that the cracks are induced by asperity contacts.

The wear rates and the micropitting propagation rates follow a similar pattern. This implies that the loss of material is mainly caused by micropitting.

Microscopic examination of the cross sections taken from tested specimens has revealed two types of microcracks: subsurface and surface initiated cracks. The direction of propagation of the subsurface cracks is parallel to the surface and their length is in the order of tens to hundreds of microns. Their direction and size suggest that they are not responsible for micropitting. Many of these parallel cracks propagate at a depth of 3-6 μm. This indicates the existence of a critical region at these depths.

The crack branching explains the coalescence of micropits. Multiple crack branches occur due to plastic deformation of the material that remains when a micropit is formed on an asperity.

Residual stress and retained austenite change their magnitude immediately after the test starts but no changes occur afterwards. This suggests that, under the conditions described in this work, residual stress and retained austenite at the surface have little effect on micropitting.

~ Chapter 6 ~

PHASE TRANSFORMATIONS

6 PHASE TRANSFORMATIONS

6.1 General

As discussed in section 3.4, research in bearings has revealed that surface contact fatigue is associated with a number of microstructural transformations grouped under the term *martensite decay*. Due to their etching characteristics, the products of martensite decay are called: *dark etching regions* (DER), *white etching bands* (WEB) and *butterflies*. The aim of this chapter is to present the results of the metallographic investigation done on disc specimens tested in eight different experimental conditions. The microstructure of the specimens has been investigated on cross sections taken in longitudinal and transverse direction after etching with nital or picral. In addition to the aforementioned phases, zones that exhibit a non-martensitic microstructure have been found near the tested surface to a depth up to 6 μm. These zones are referred to as *plastic deformation layer* (PDL). The mechanical properties (i.e., hardness and elastic modulus) of these phases have been determined by the nanoindentation technique.

6.2 Martensite decay

The microscopic examination of the samples etched with nital has revealed similar microstructural features to those reported in fatigued bearings, namely *dark etching regions* (DER) and *white etching bands* (WEB). Figure 6-1 shows two cross sections of the disc tested in experiment 1 taken in longitudinal direction in (a) and, transverse direction in (b). The dark etching regions are clearly seen in both pictures, beneath the specimen surface. These regions appear much darker than the rest of the microstructure. Beneath the dark etching zone, the microstructure is characterised by a preferential oriented texture (see Figure 6-1a). The texture consists of white lines (white etching bands-WEB), which lie between 185 and 690 μm below the contact surface (marked on the picture). In Figure 6-1a the white etching bands are inclined to the surface plane to

an angle $\alpha = 45°$ in the direction opposite to the sliding direction. WEB are not observable as inclined bands in the transverse section from Figure 6-1b but as a zone of fine microstructure, lying beneath the DER, slightly darker than the unaffected microstructure, which extends to a depth approximately equals to the depth from Figure 6-1a.

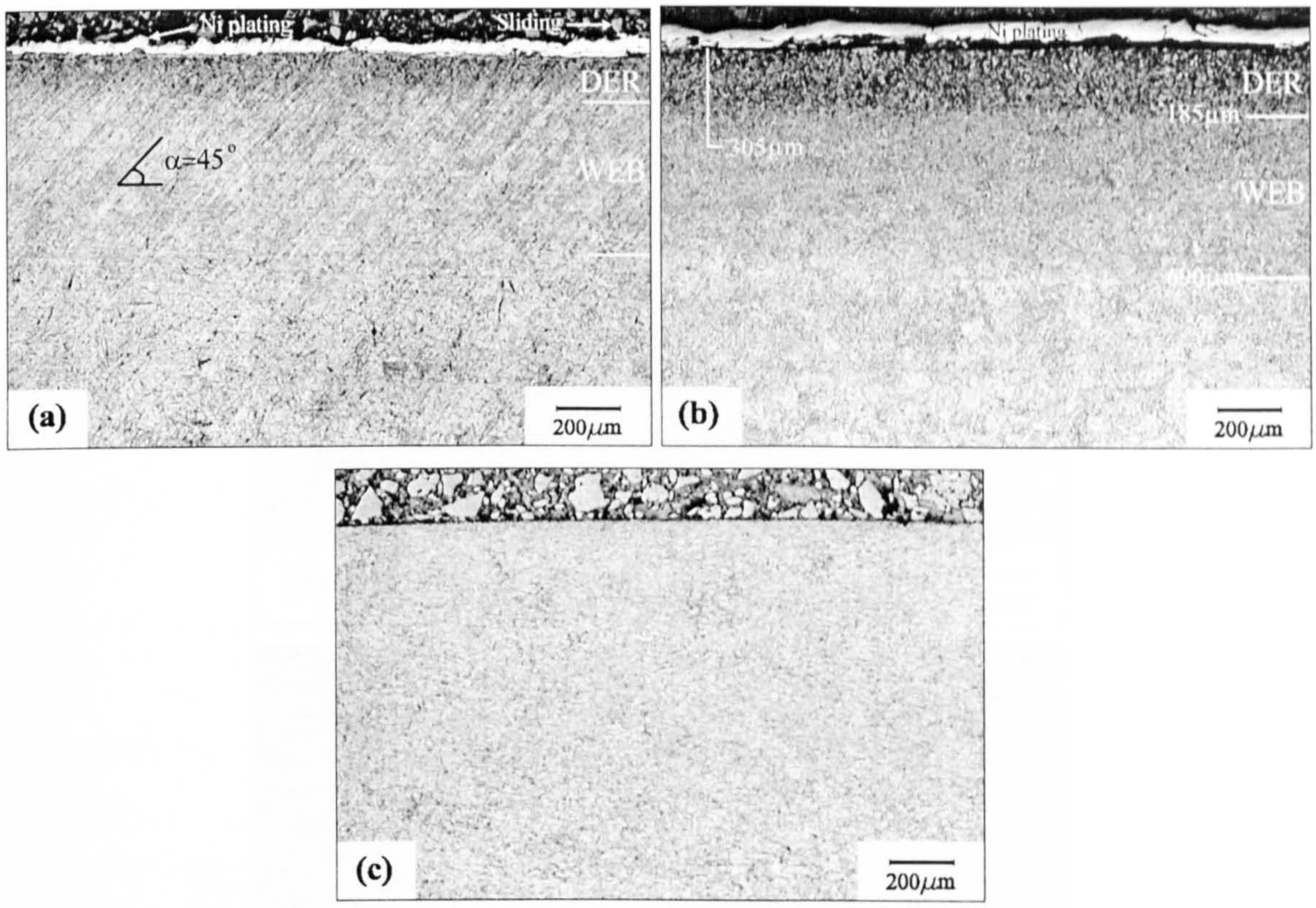


Figure 6-1. Dark etching regions and white etching bands developed after test. Experiment 1; Etch: nital 2%. (a) Longitudinal section. (b) Transverse section. (c) Longitudinal section before test showing no DER or WEB.

Figure 6-2 shows the microstructure of the specimen tested in experiment 7. The development of DER and WEB phases is more advanced as compared to the previous case although both specimens were tested for the same number of stress cycles. As marked in Figure 6-2b, DER extend to a depth of 140 μm and WEB to a depth of 870 μm. Since in experiment 7 the contact pressure and the temperature were higher it can be assumed that higher loads and temperatures are favourable to the martensite decay phenomenon. In the longitudinal direction (Figure 6-2a), the dark etching regions extend to higher depths inside the WEB region and appear between the white bands (see

the arrow). The angle of inclination of the white etching bands, α =21° is shallower than in the previous case. This difference should be related to the magnitude of sliding. In experiment 1 speed and slide-to-roll ratio are both at the highest level while in experiment 7 they are at the low level. It results that the angle of inclination of the white etching bands increases with speed and slide-to-roll ratio. The examination of the transverse section reveals two sets of white etching bands having opposite orientations and inclined relative to the surface plane at $\alpha_1 = 55°$ and $\alpha_2 = 38°$, respectively.

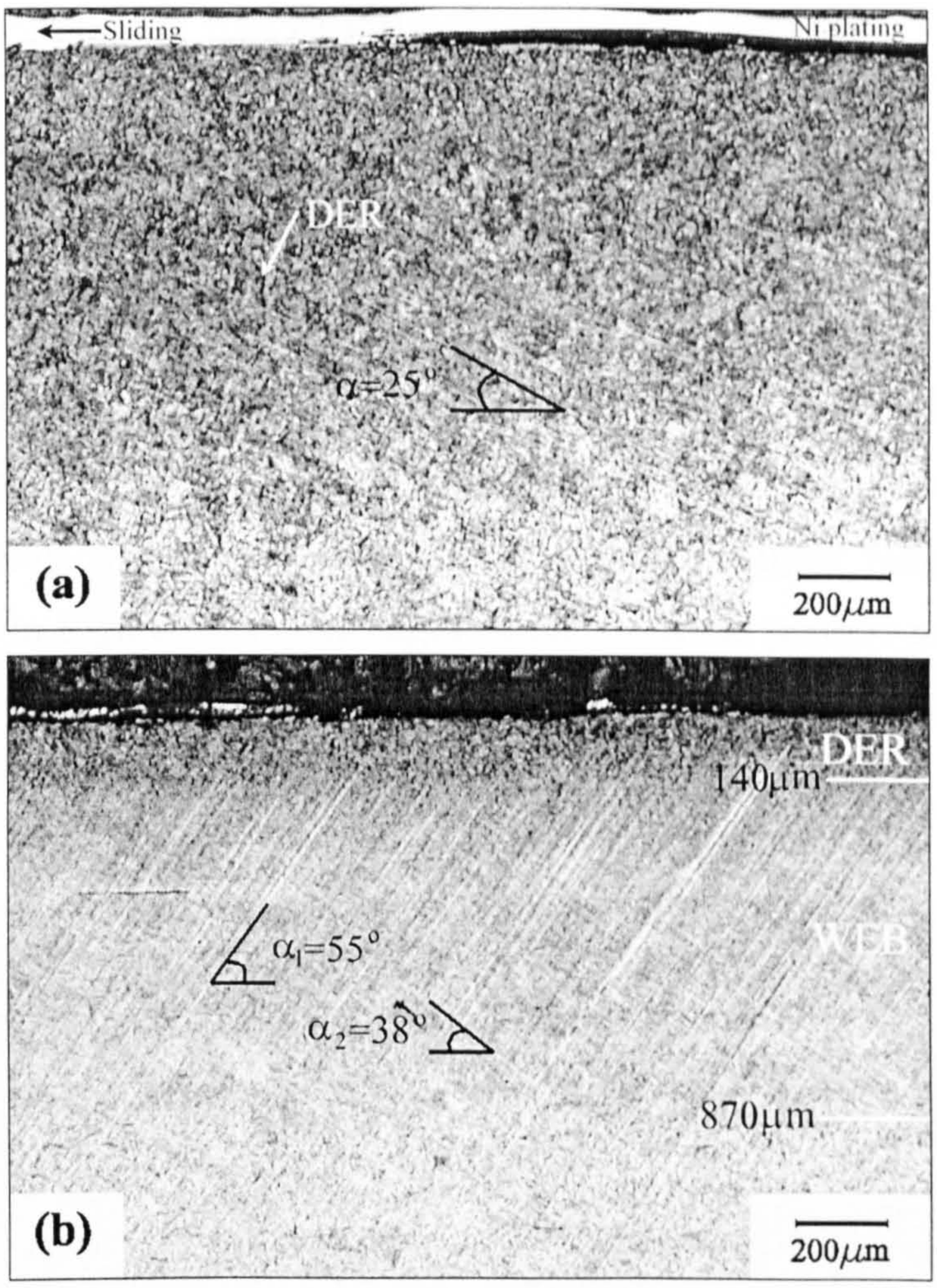


Figure 6-2. Dark etching regions and white etching bands. (a) Longitudinal section. (b) Transverse section. Experiment 7; Etch: nital 2%.

The two sets of WEB suggest that the transformation is dependent on the Hertzian stress conditions and the values of the angles, α_1 and α_2, suggest that the texture develops in the direction of the principal stresses. Moreover, the transformed region (DER + WEB) is centred to a depth that corresponds to the depth of maximum shear stress determined by the Hertzian theory. Table 6-1 lists the depth of maximum shear stress, z, and the

centre of the region with decayed martensite (the half-width of the region DER + WEB) for the specimens tested in experiments 1, 6 and 7, the only where WEB has been observed.

Table 6-1. The correlation between the depth of maximum shear stress and the centre of the transformed region

Experiment	Depth of maximum shear stress z (μm)	Centre of the transformed region (DER +WEB) (μm)
1	336.5	345
6	425.5	410
7	430	435

The microstructures from Figure 6-3 have been revealed by picral etch. Coarse carbide particles can be seen in (a) in the region beneath the surface, corresponding to the depth of martensite decay. The picture from Figure 6-3b is a higher magnification of the particle arrowed in Figure 6-3a. The edges of the spherical particle are heavily etched by picral which attacks the carbide boundaries. It can be concluded that the formation of DER and WEB is associated with a process of carbide precipitation.

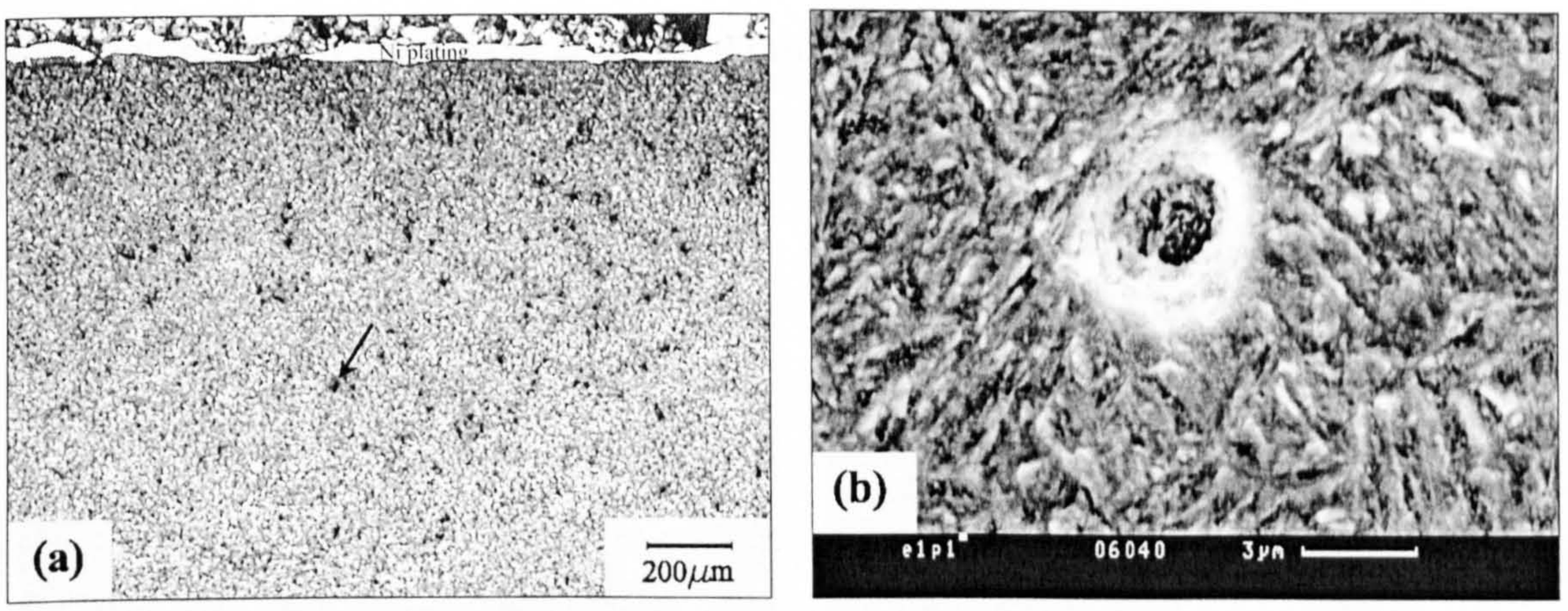


Figure 6-3. Carbide precipitation. (a) Light microscopy image. (b) Secondary electron image. Experiment 6; Etch: picral.

In the following sections, a more detailed examination of the products of martensite decay will be presented.

6.2.1 Dark etching regions

The SEM images from Figure 6-4 were taken from a region at about 10 μm below the surface of the sample tested in experiment 6. The micrograph from Figure 6-4a shows a circular distribution of the dark etching phase grains. The diameter of the region is in the range of the diameter of the prior austenite grain, which means that the DER phase initiates at the prior austenite grain boundaries. The grains grow and break down the martensite needles until the entire martensitic microstructure, within the prior austenite grain, is transformed (see Figure 6-5b).

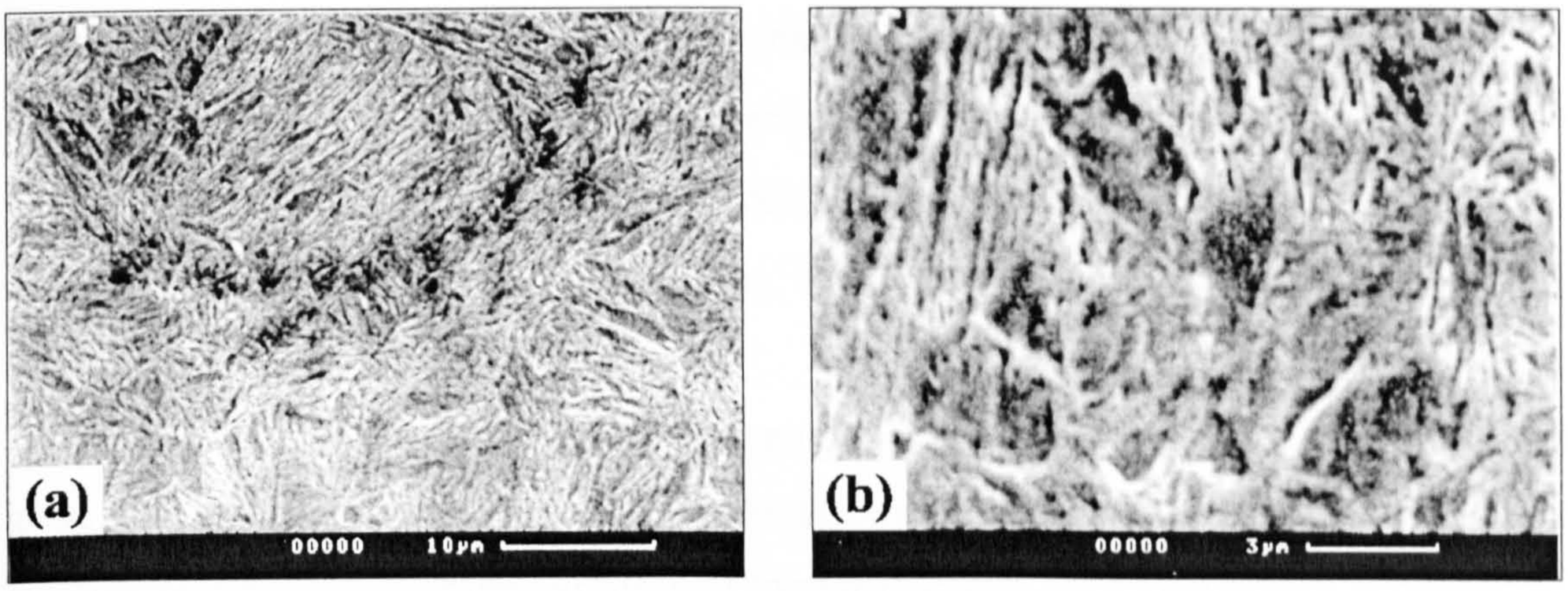


Figure 6-4. Nucleation of dark etching regions. Secondary electron images. (a) DER grains at the prior austenite grain boundary. (b) Higher magnification of DER. Experiment 6; Etch: nital.

The picture from Figure 6-5a shows the microstructure in a region where DER has not been initiated.

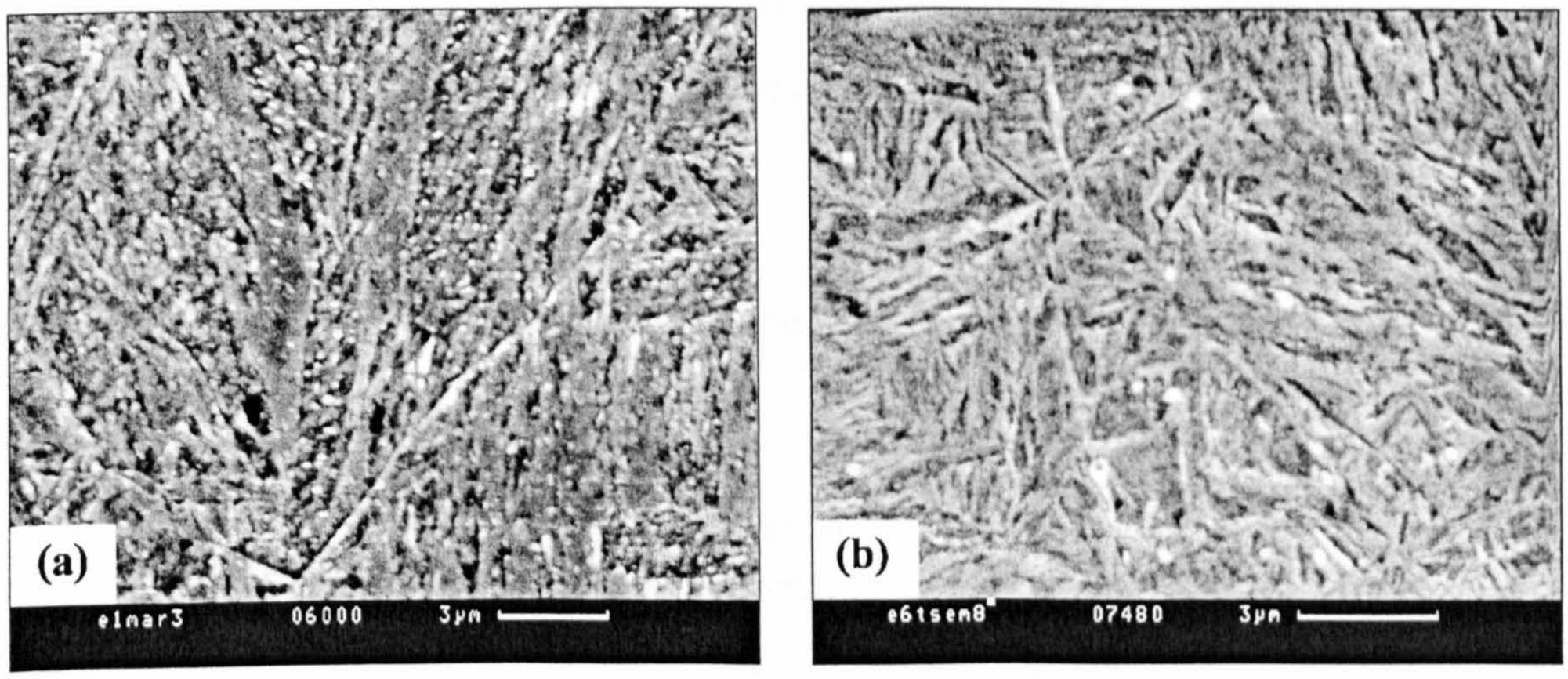


Figure 6-5. Carbide precipitation. Secondary electron images. Experiment 6. (a) Picral etch. (b) Nital etch.

The prior austenite grain boundary is intact and, numerous, small, spherical carbides can be seen within the martensite needles. In Figure 6-5b the DER grains appear to break down the martensite needles in the prior austenite grain boundary and the number of carbide precipitates is reduced. The coarse carbides appear at the boundaries between the initial martensite and the newly formed DER grains.

The precipitation of carbides, suggests that temperature plays a significant role in the development of the martensite decay products by facilitating the diffusion of carbon.

In order to assess the mechanical properties of DER phase a nanoindentation test has been carried out on the specimen tested in experiment 1. 40 indentations, arranged in the array shown in Figure 6-6 have been performed, between 35 and 305 μm below the surface, in the location marked in the upper left corner of Figure 6-1b after re-polishing the specimen using 1 μm diamond paste. The hardness values corresponding to each indentation are recorded in Table 6-2.

It can be seen that there is a very little scattering in the data corresponding to the DER zone compare to the WEB zone. This indicates that DER has a fine, homogeneous microstructure. The average hardness of the DER is $H = 12.3$ GPa which is higher than the hardness of the original martensite (i.e. $H \approx 10$ GPa). The large scattering in the data obtained from the WEB zone, from as low as $H = 11.7$ GPa to $H = 22$ GPa, is due to the different properties of the white bands and the space between them. These values indicate that the mechanical properties of the white etching bands are similar to those of the original martensite and they are separated by regions where a hardening process has taken place. The presence of the coarse carbides suggests a process of precipitation hardening.

More detailed investigations on the WEB are presented in the following section.

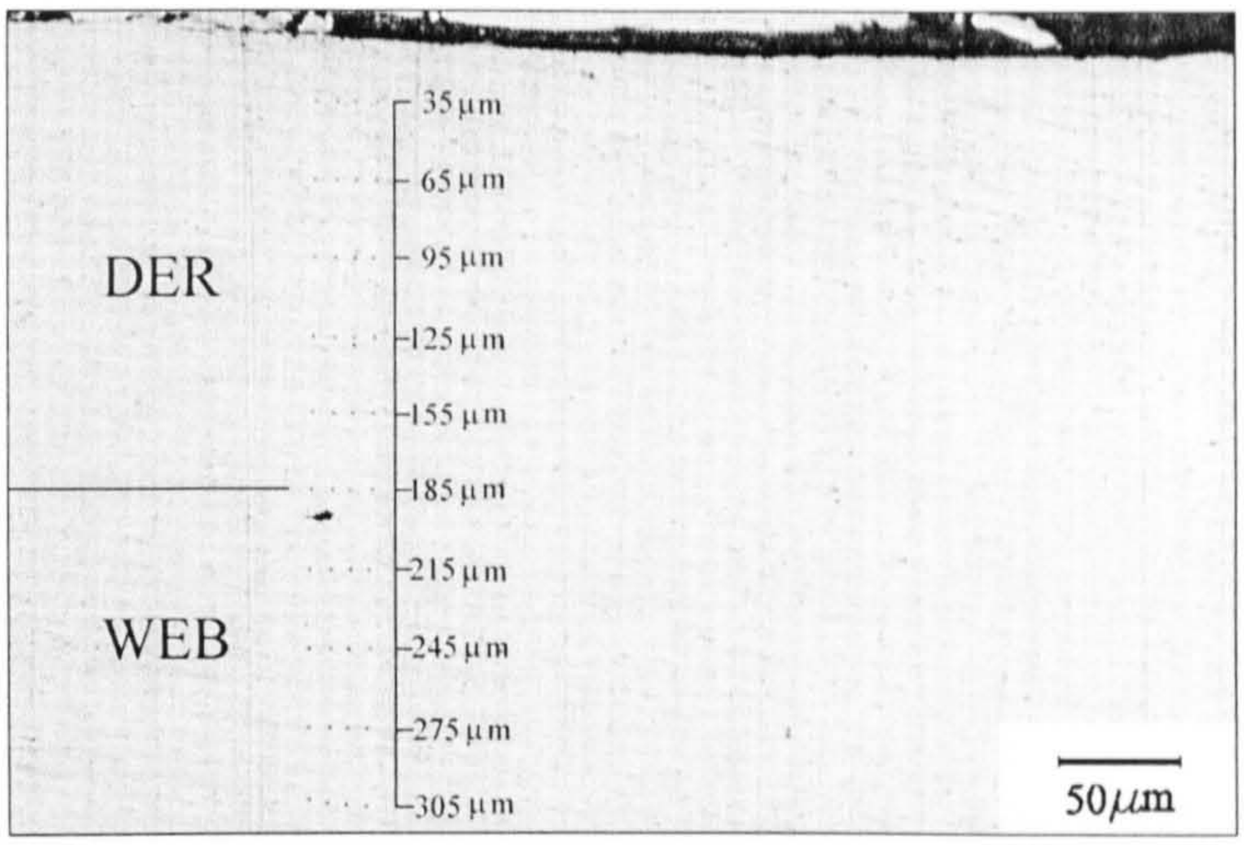


Figure 6-6. The nanoindentation array used to asses the mechanical properties of DER.

Table 6-2. Nanoindentation data corresponding to the array from Figure 6-6.

Depth (µm)	Hardness, *H* (GPa)			
35	13.4	12.9	12.6	12.6
65	13.3	12.6	12.4	13.5
95	12.8	12.3	12.3	12.3
125	12.0	11.2	12.5	11.8
155	11.9	11.9	11.4	12.7
185	11.8	12.6	12.4	11.7
215	12.9	13.9	13.5	13.1
245	14.4	15.9	12.3	13.8
275	15.6	15.9	18.5	19.8
305	-	18.0	21.6	22.0

6.2.2 White etching bands

In Figure 6-7 a certain directionality of the microstructure is distinguishable. The directionality consists of interrupted bands that are inclined to the surface to an angle $\alpha = 61°$ in the direction opposite to sliding. This pattern is similar to the white etching bands observed in Figures 6-1a and 6-2a. The bands from Figure 6-7 are in the early stages of development and they have thin dark boundaries, which, probably consists of carbide precipitates. This arrangement is consistent with the 30° bands described by Swahn et al (1976).

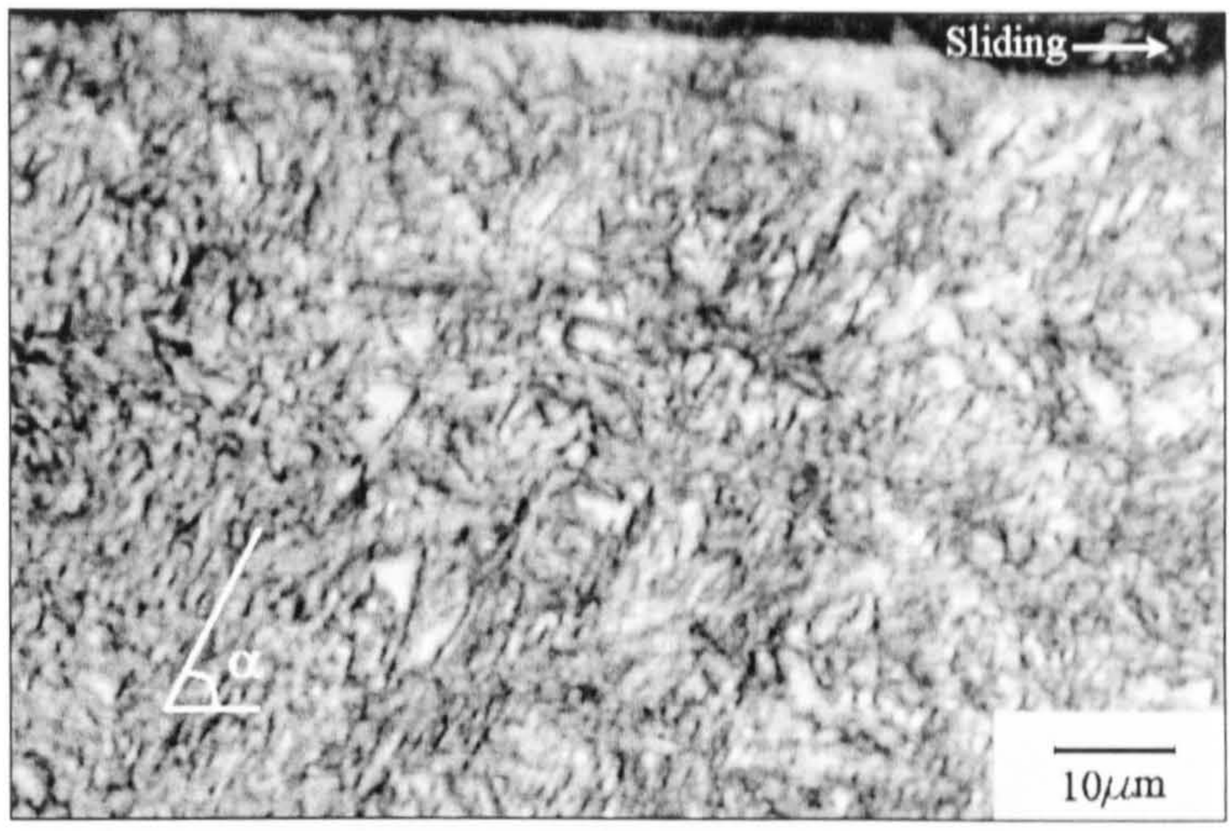


Figure 6-7. White etching bands in the initiation stage. Experiment 4; Etch: nital.

The micrograph from Figure 6-8 is a higher magnification of the WEB region shown in Figure 6-1a. The martensitic structure is completely replaced by white etching bands of few microns width, separated by larger zones of about 10 μm. The line marked with the letter B is a polishing artefact. Very thin bands (in A), perpendicular to the white etching bands and crossing over them are believed to be non-transformed martensite needles.

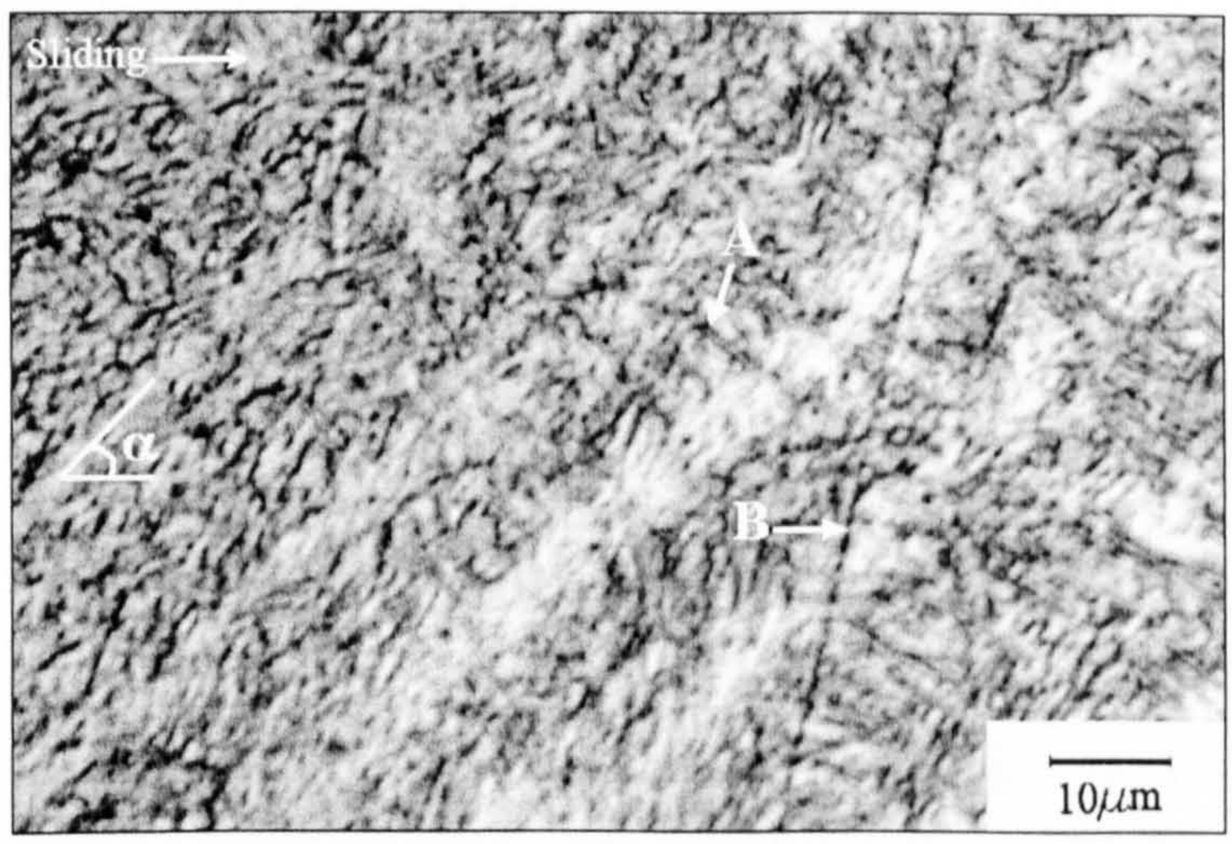


Figure 6-8. White etching bands. Experiment 1; Etch: nital.

Figure 6-9 depicts a high magnification of the space between the white etching bands. The microstructure is abundant in carbide precipitates, which explain the high hardness recorded by nanoindentation in this region.

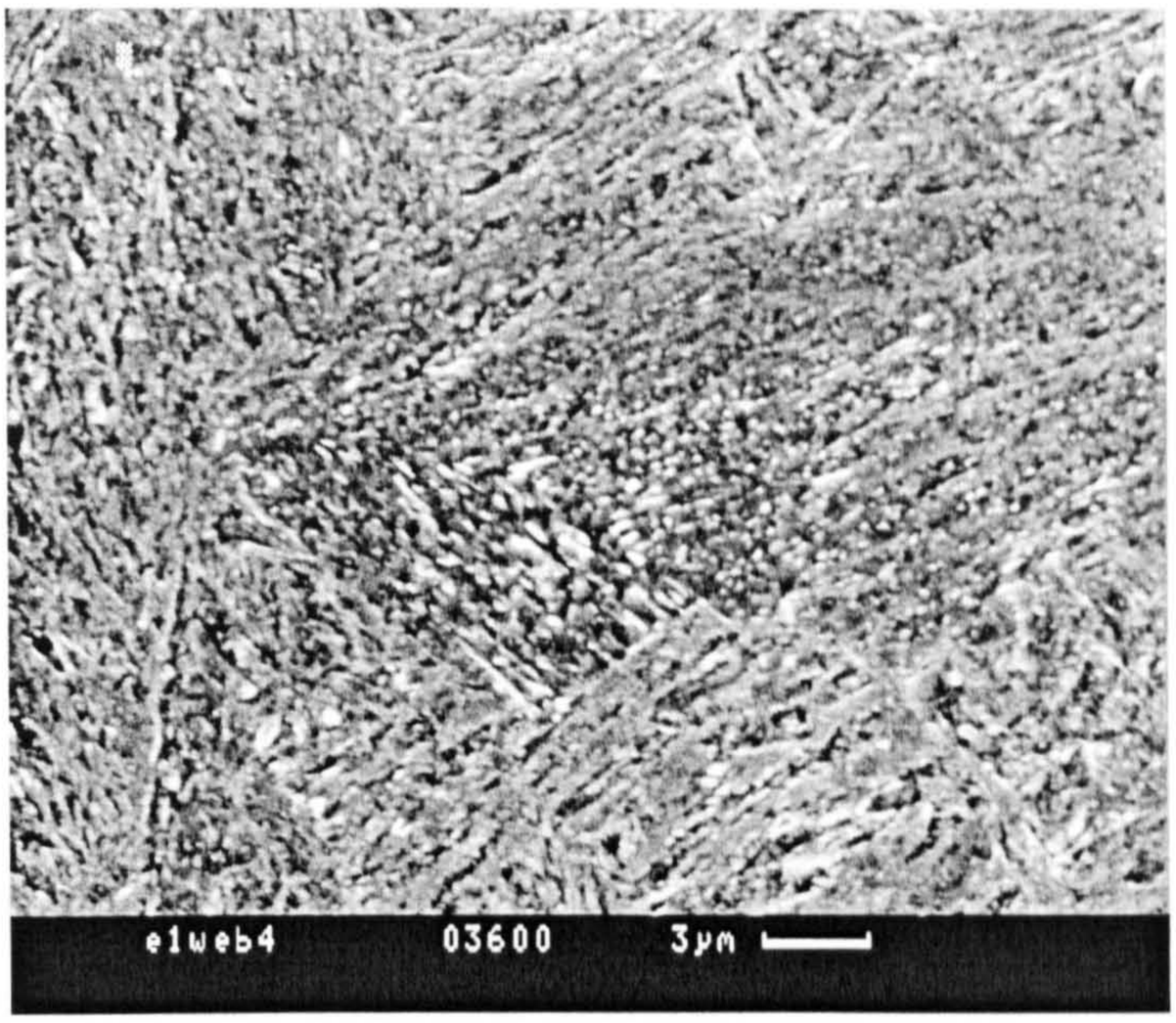


Figure 6-9. Space between the bands. Secondary electron image. Experiment 1; Etch: picral.

In order to determine the hardness of the bands and the space between the bands the specimen shown in Figure 6-1a has been slightly re-polished to reduce the effect of roughness but at the same time to allow the observation of the microstructure. The indentation array is shown in Figure 6-10.

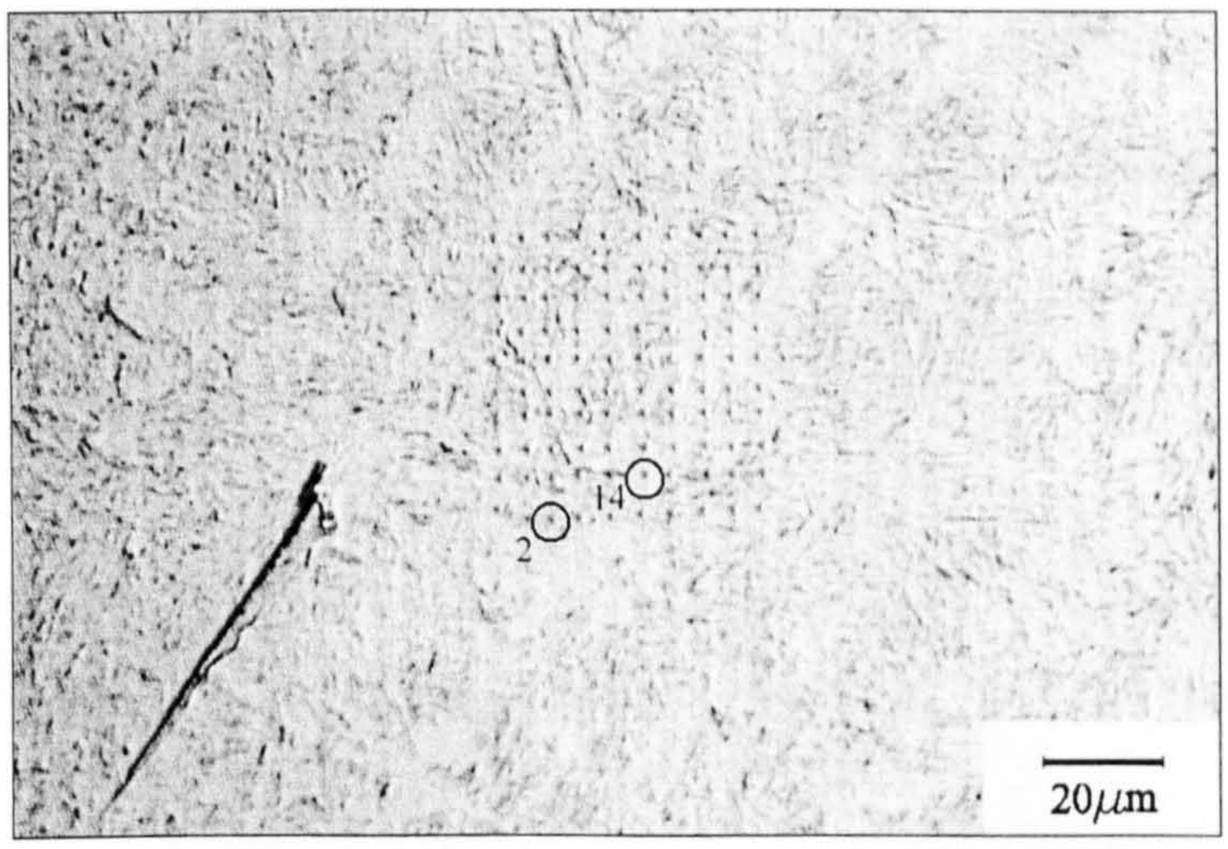


Figure 6-10. Nanoindentation array used to asses the mechanical properties of WEB.

Table 6-3. Nanoindentation data corresponding to indents 2 and 14.

Indentation	*E* (GPa)	*H* (GPa)
2	157.3	20.3
14	196.4	11.6

The indentations number 2 and number 14 (circled in Figure 6-10) have been performed in the space between the bands and in the band itself, respectively. The recorded elastic modulus, E, and the hardness, H, are given in Table 6-3. These results are in good agreement with the data corresponding to WEB in Table 6-2. The white etching bands are slightly harder than the original martensite but the space between the bands is considerably harder. The hardening that occurred between the bands is explained by the intense carbide precipitation observed in Figure 6-9.

The AFM images from Figure 6-11 show that the surface roughness has been considerably reduced after re-polishing so the nanoindentation results are reliable. The pile-up phenomenon is also observed, therefore the use of the correction factors determined by Oila & Bull (2003) is justified.

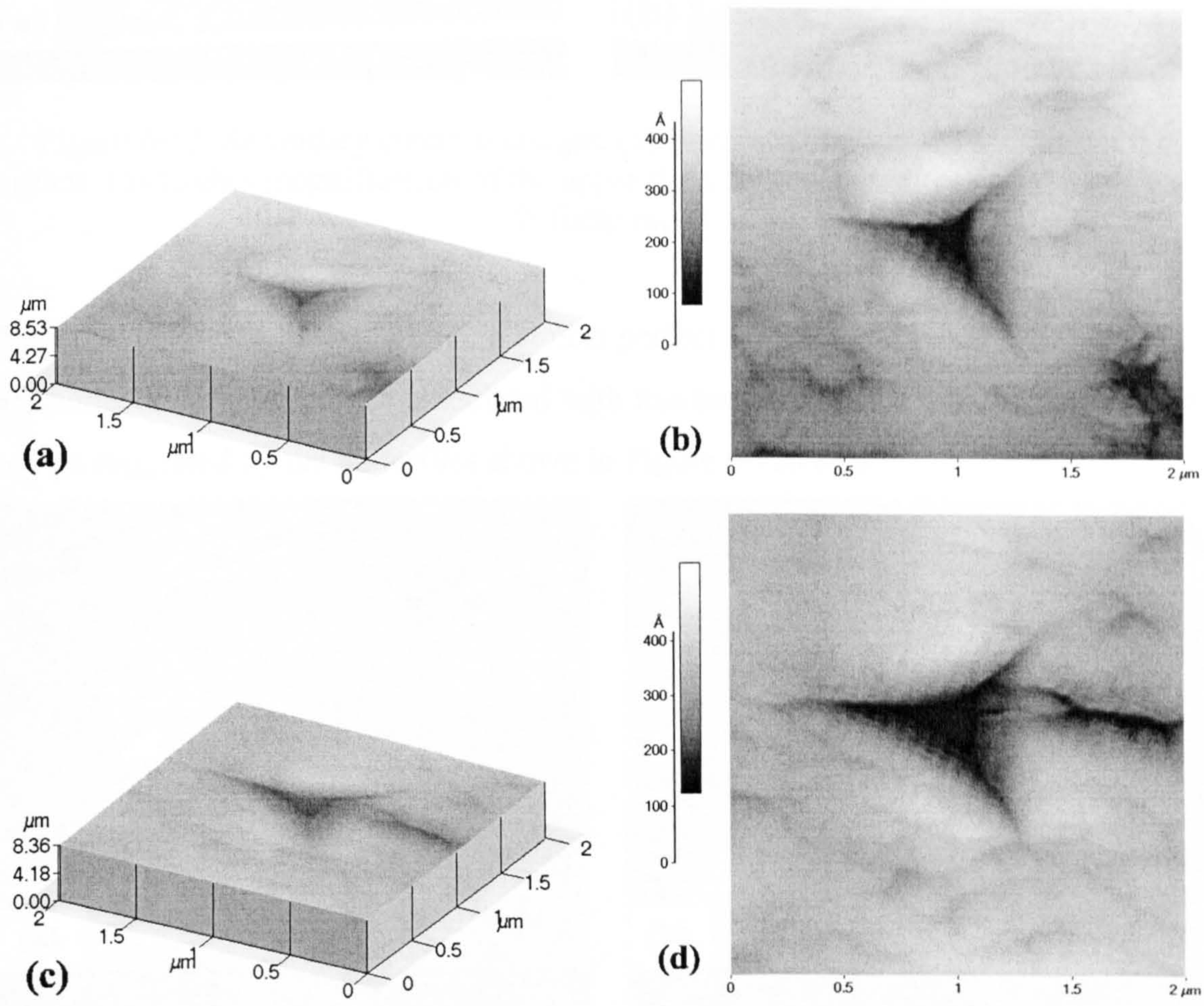


Figure 6-11. AFM images of the indentations number 2 in (a) & (b) and 14 in (c) & (d).

6.2.3 Butterflies

A third microstructural feature present in fatigued specimens is represented by the butterflies. Figure 6-12a shows the microstructure of the specimen tested in experiment 7 in the region near the surface. The upper right corner of Figure 6-12a is further magnified in Figure 6-12b.

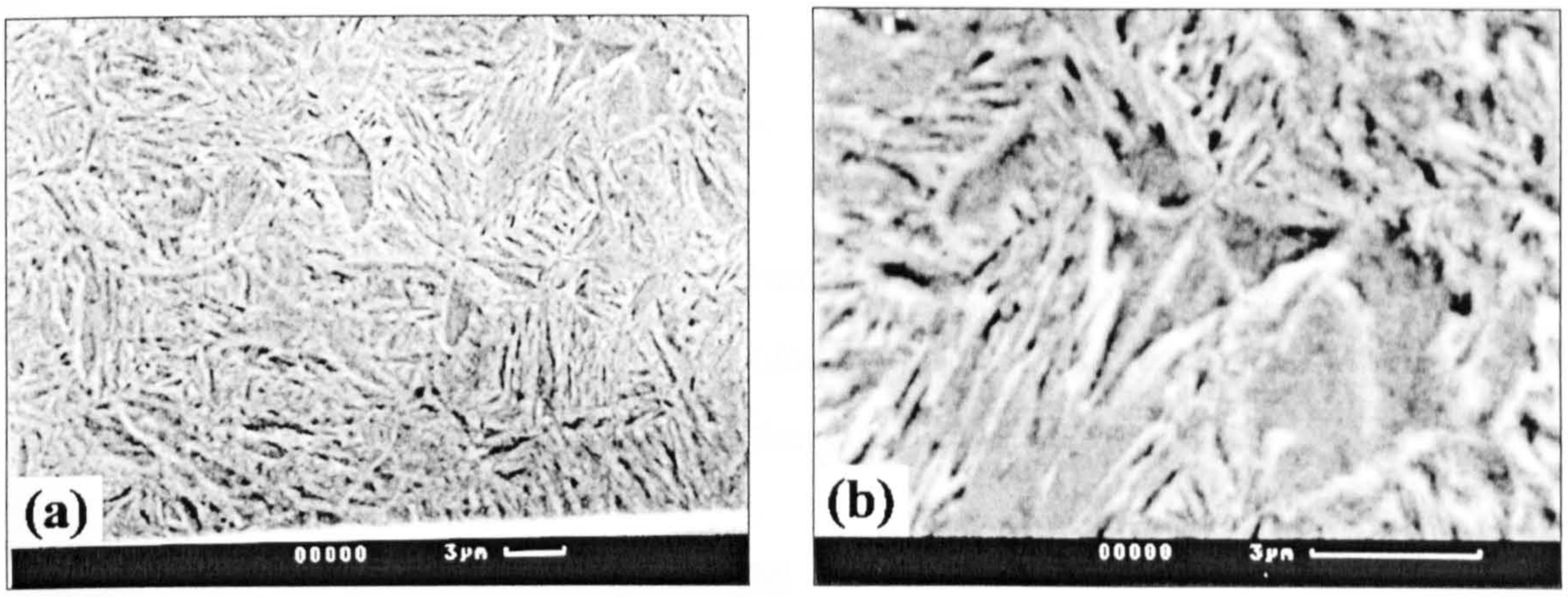


Figure 6-12. Secondary electron images. (a) Microstructure in the vicinity of the surface. (b) Higher magnification of the upper right corner region from (a). Experiment 7; Etch: nital

The picture shows a butterfly having almost perfect symmetrical wings. No observable inclusions or carbides can be associated with this butterfly. The shape of the wings can vary as suggested by the butterflies shown in Figure 6-13a & b.

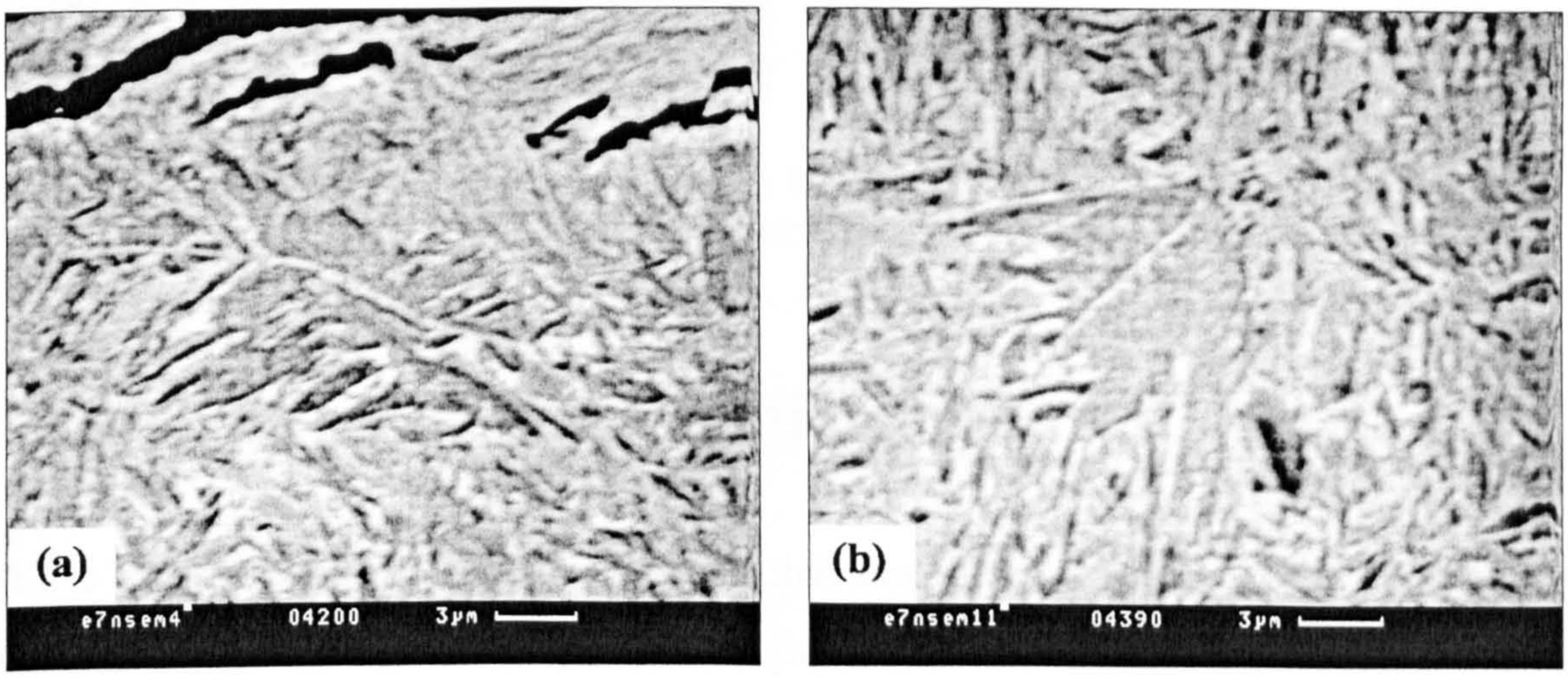


Figure 6-13. Butterflies formed near the surface of the specimen tested in experiment 7. Etch: nital.

Their shape can vary but the wings appear to develop symmetrically to a central axis. Since non-metallic inclusions or carbide particles cannot be observed within these butterflies but the wings develop symmetrically to a central axis it is suggested that the axis of symmetry is the prior austenite grain boundary and the wings represent zones of DER. The axis of symmetry is not related to the direction of sliding. The butterflies shown in Figure 6-13 were both found in the same specimen and they have opposite orientations. This supports the aforementioned observation that the central axis represents the prior austenite grain boundary.

The nanohardness value, H=12.5 GPa determined for the butterfly shown in Figure 6-14a confirms the above hypothesis and also the mechanism of butterfly formation by recrystallisation and work hardening of the newly formed grains, described in section 3.4.2.2.4.

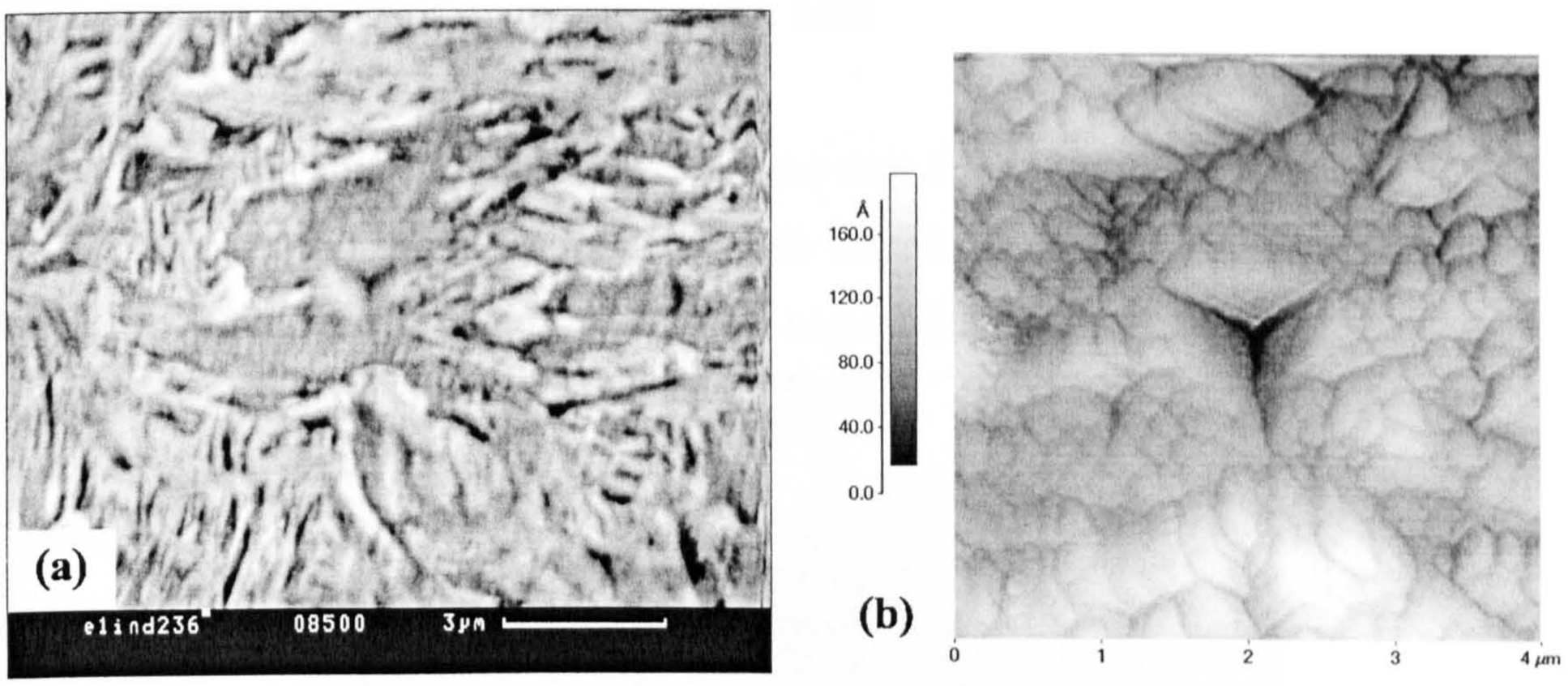


Figure 6-14. (a) SEM image of a butterfly and the indentation. (b) AFM image of the indentation. Experiment 1. Etch: nital.

6.3 Plastic deformation layer

It has been shown in the previous chapter that asperities are subjected to severe plastic deformation. The regions very close to the surface, underneath asperities are also subjected to severe plastic deformation. This is illustrated by the micrograph from Figure 6-15 in which the microstructure appears deformed in the direction of sliding. The crack appears to be initiated at the surface and its propagation follows the direction

of deformation. It is suggested that the accumulated plastic deformation leads to the formation of the white layer shown in Figure 6-16. This microstructural feature will be referred in this work as the *plastic deformation layer* (PDL).

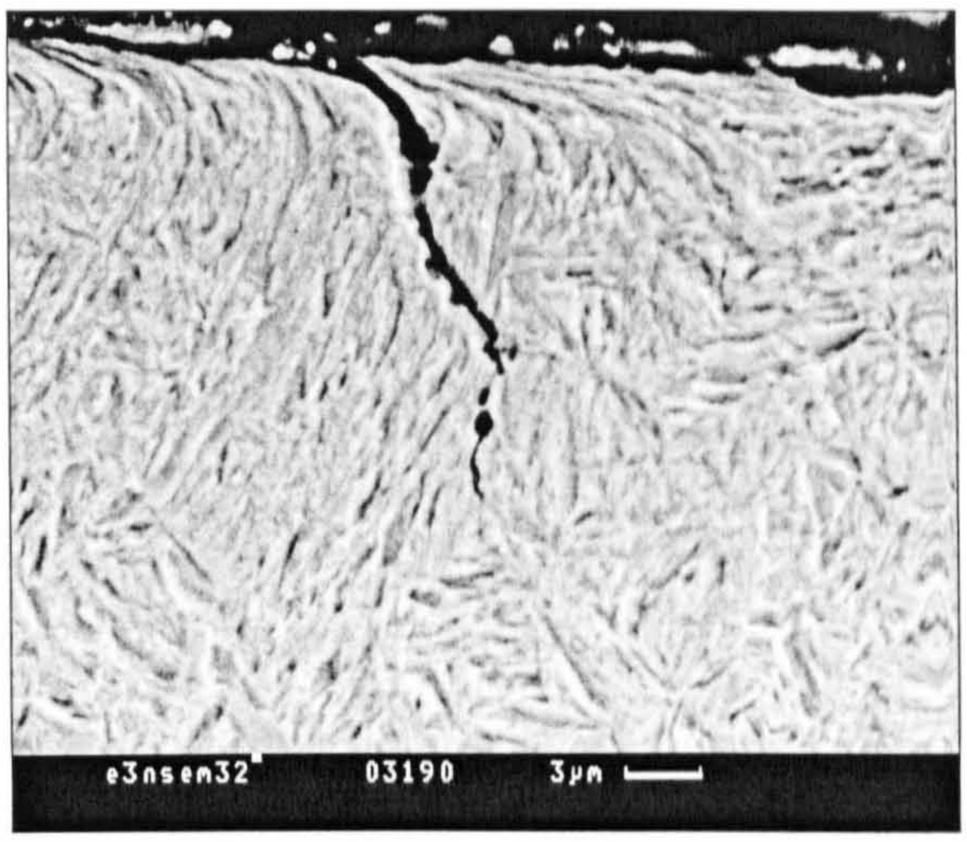


Figure 6-15. Plastic deformation near the surface. Experiment 3, longitudinal section; Etch: nital.

As the picture from Figure 6-16 indicates, the first micropits form within the PDL. The PDL exhibits an apparently non-martensitic microstructure (the absence of needles). The micrograph from Figure 6-17a shows a higher magnification of the PDL.

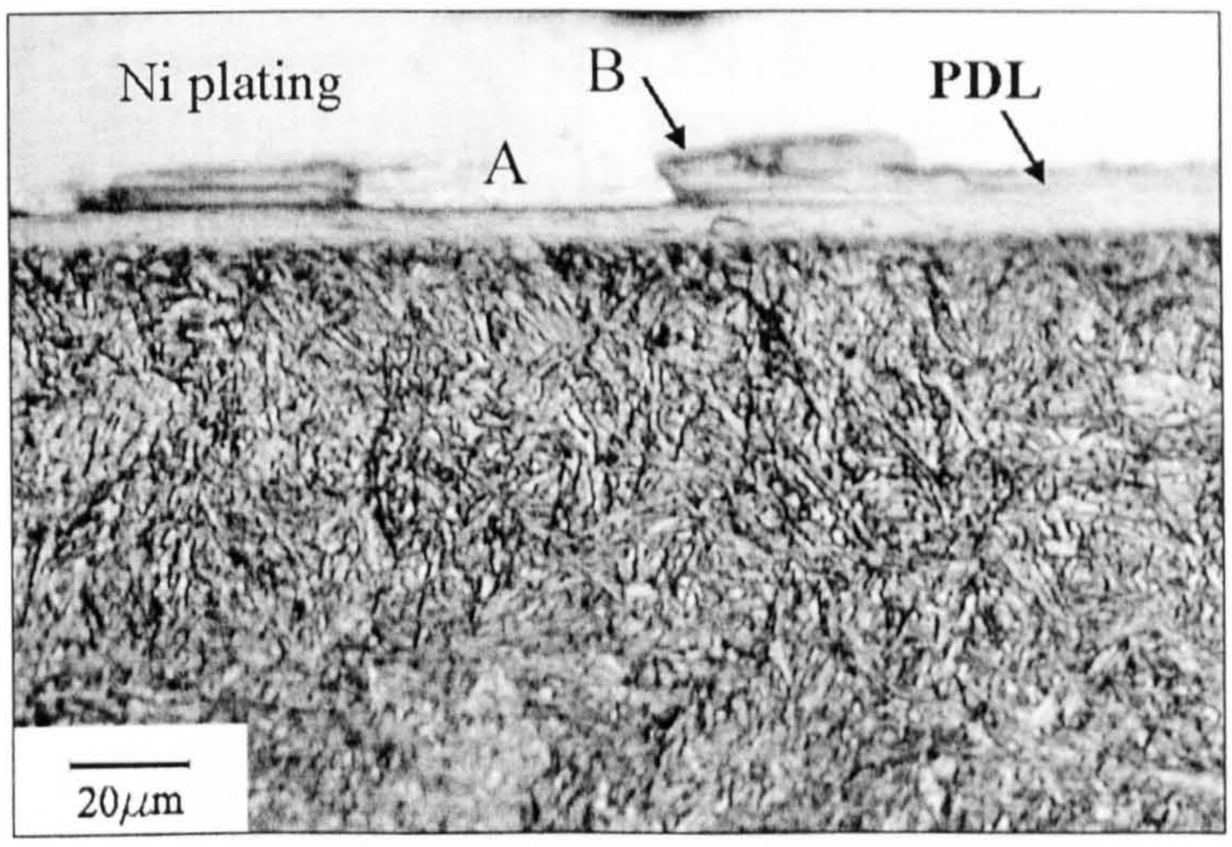


Figure 6-16. Plastic deformation layer. Light microscope image. Experiment 7, transverse section; Etch: nital.

Nanoindentation tests have been performed in the region marked B in Figure 6-16. This region and the indentation array are shown at higher magnification in Figure 6-17b. The nanoindentation data are given in Table 6-4 and they suggest that the PDL

microstructure undergoes a work hardening process. It follows that, near the surface of the fatigued specimens, a high hardness layer occurs. Big DER grains can be seen in Figure 6-18a at the boundary of the PDL. The PDL does not always appear as a continuous layer but most often, as semi-circular zones, usually less than 10 μm diameter (see Figure 6-18b). The shape and the size of the PDL zones suggest they are formed due to the stress concentration induced by asperities.

The metallographic examination indicates that the boundaries of these regions are preferential sites for microcrack initiation and propagation. The depth to which the plastic deformation layer extends coincides with the depth of the parallel cracks. It can be assumed that the critical region where the parallel cracks initiate and propagate (see section 5.7.1) is the boundary of the PDL.

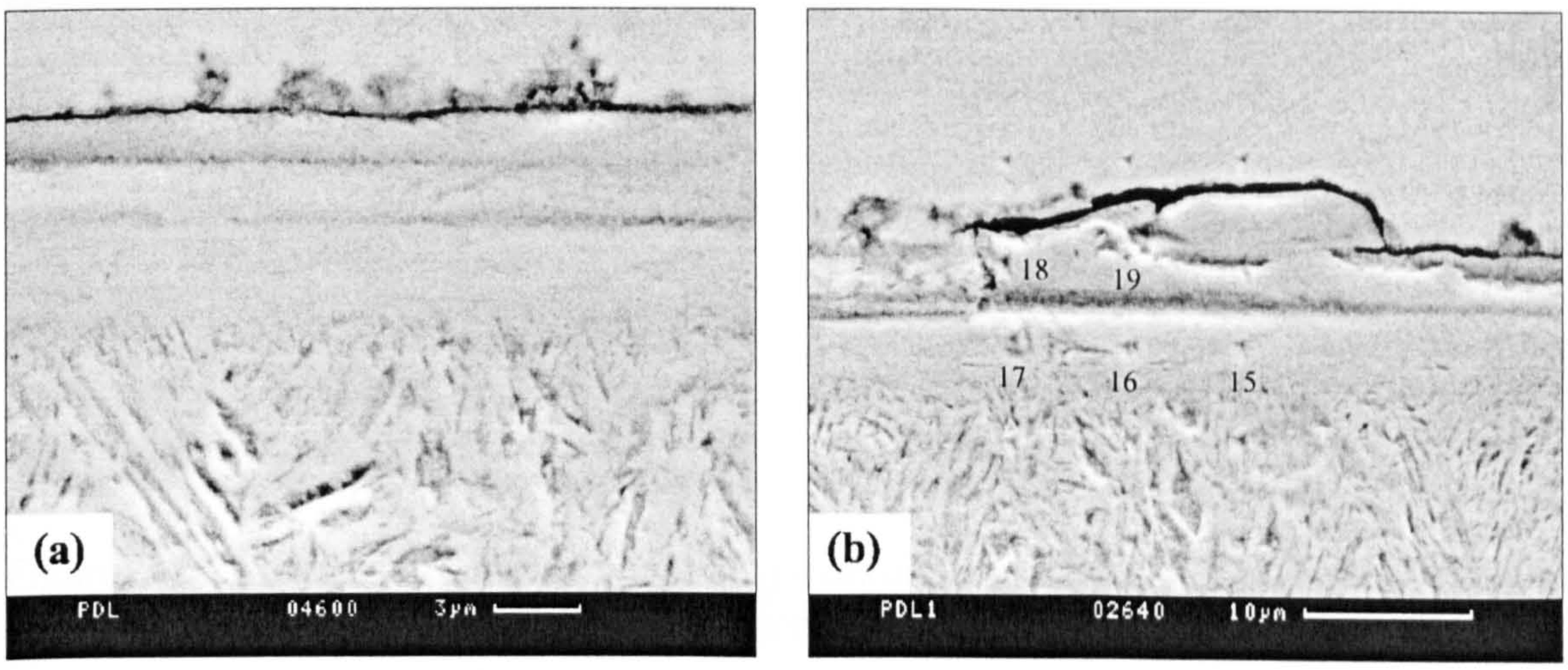


Figure 6-17. (a) Plastic deformation layer. (b) Nanoindentation array for testing PDL (region marked B in Figure 6-16).

Table 6-4. Nanoindentation data corresponding to PDL.

Indentation	*E* (GPa)	*H* (GPa)
15	180.6	12.8
16	190.2	14.1
17	196.1	15.2
18	149.5	8.5
19	184.5	13.0

The aspects seen on the micrographs from Figure 6-19 are just two examples of the most often observed feature in fatigued specimens. The surface initiated cracks

propagate at the boundary of the plastic deformation region. It appears that the two types of cracks, i.e., parallel and inclined to the surface, are most commonly associated with two types, geometrically different, of the plastic deformation region. If the plastic deformation region has developed as a continuous layer then the cracks initiate and propagate along the layer's boundary, parallel to the surface. If the plastic deformation region has developed as a semi-circular zone beneath an asperity the crack initiates at the surface, at the intersection point between the surface and the boundary and propagates along the boundary, in a direction inclined to the surface. No inclined cracks have been observed to propagate through the PDL.

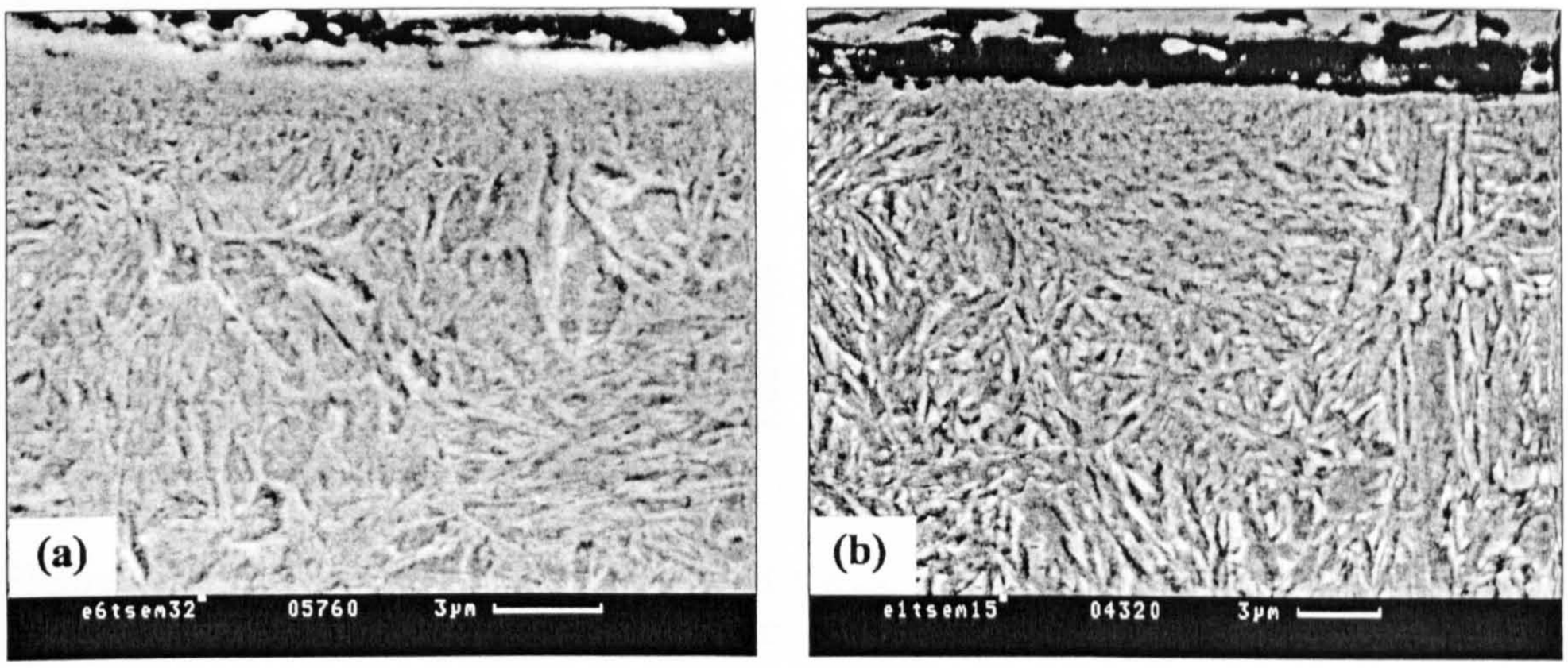


Figure 6-18. Plastic deformation region. (a) Continuous layer. (b) Semi-circular region. Experiment 1, transverse section; longitudinal lay. Etch: nital.

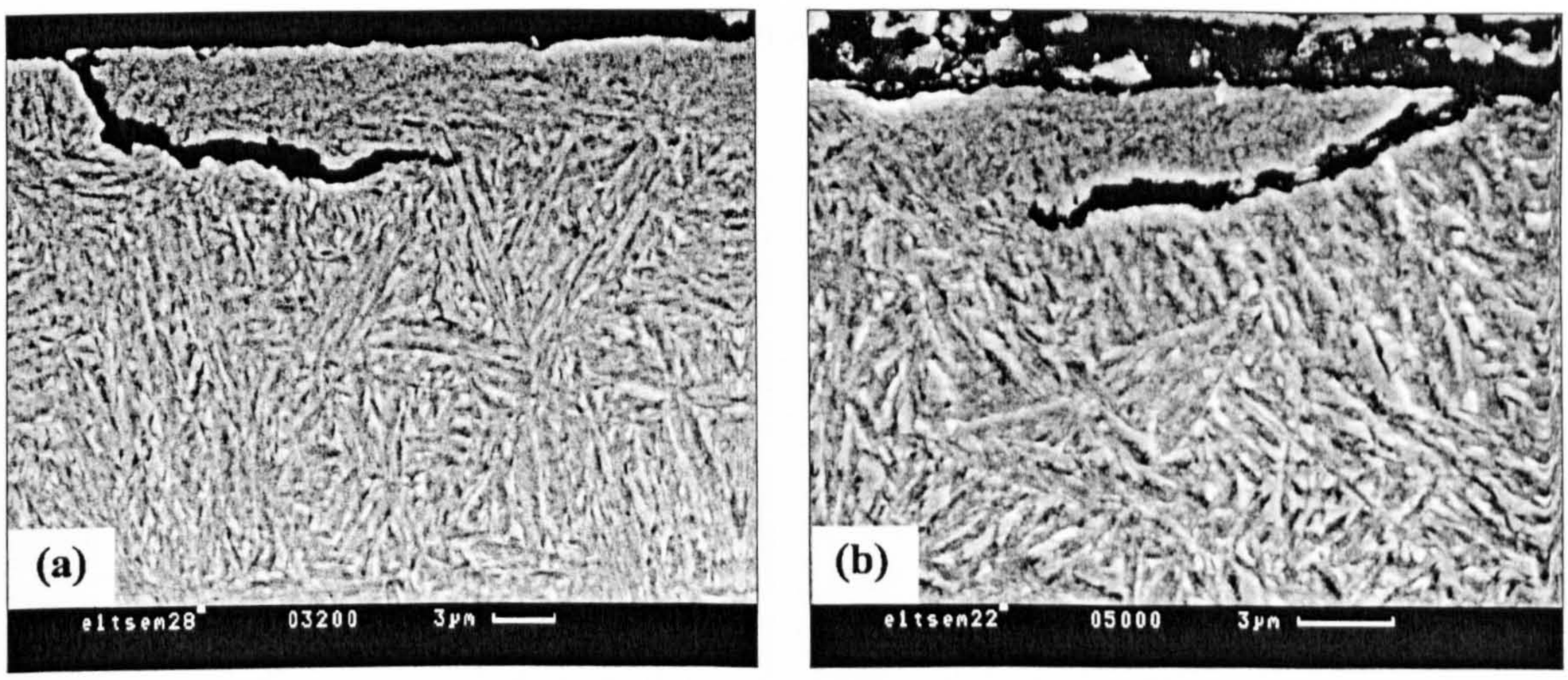


Figure 6-19. SEM images showing crack propagation at the PDL boundary. Experiment 1, transverse section; longitudinal lay. Etch: nital.

The picture from Figure 6-20a shows a wear debris particle detached from the specimen surface by the mechanism described above. A nanoindentation test has been performed on this wear particle according to the array shown in Figure 6-20b. The first line of indentations (i.e, from 1 to 4) is to close to the edge of the wear particle therefore these data were omitted from calculations. The hardness values obtained from the indentations 5 to 14 were averaged and the mean value is $H = 5.95$ GPa. This indicates that the material between the crack and the specimen surface is very soft compared to the surrounding material.

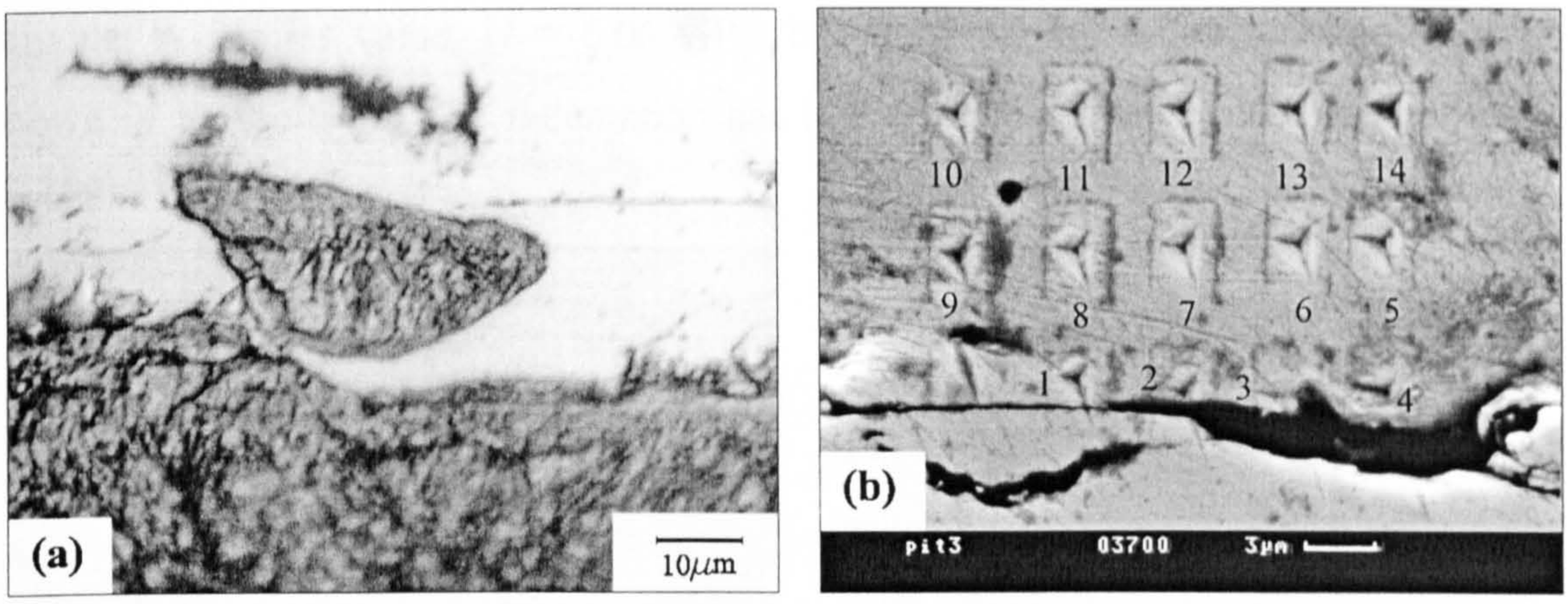


Figure 6-20. (a) Light microscope image of a wear debris. Nital etch. (b) SEM image of the indentation array on the wear debris.

Precautions in interpreting the nanoindentation results must be taken because the thickness of the wear particle is not known and, if the total displacement is over 10% (Pharr & Oliver, 1992) of the thickness, then the data are influenced by the substrate (mounting resin in this case).

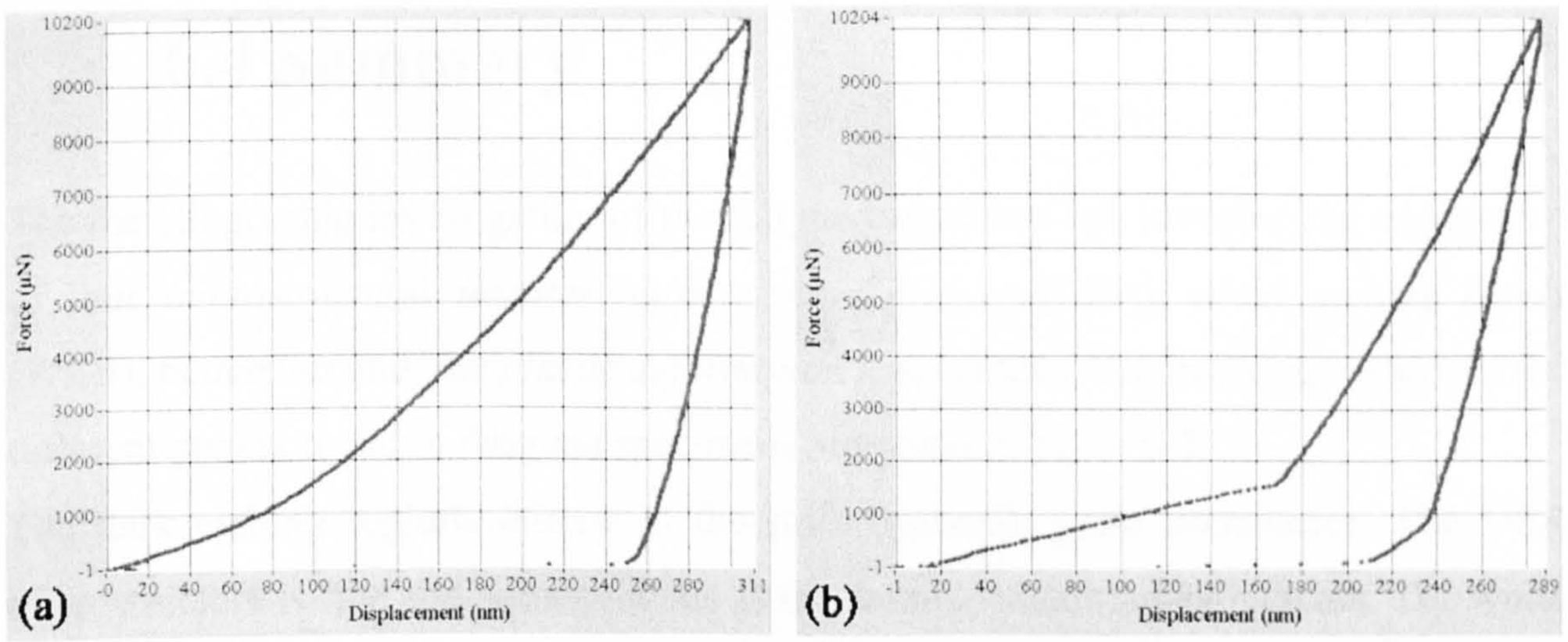


Figure 6-21. Load displacement curves for indentations 14 in (a) and 3 in (b).

The influence of the substrate can be observed on the load-displacement curves. Figure 6-21 shows the load-displacement curves for indentation 14 in (a) and 3 in (b).

In Figure 6-21b, the abrupt change in the slopes of the loading and unloading curves indicates the indentation of two different materials, at first, a soft material (the mounting resin) and then a hard one (the steel). In this particular case the edges of the wear particle are sunk in the mounting resin similar to an iceberg. The load-displacement curves for indentations from 5 to 14 appear as that from Figure 6-21a, in which the substrate does not influence the results. Therefore, the hardness value, H = 5.95 GPa is reliable. A similar value, H = 6.06 GPa, has been determined from the indentation shown in Figure 6-22. The indentation has been performed in the PDL zone near the crack.

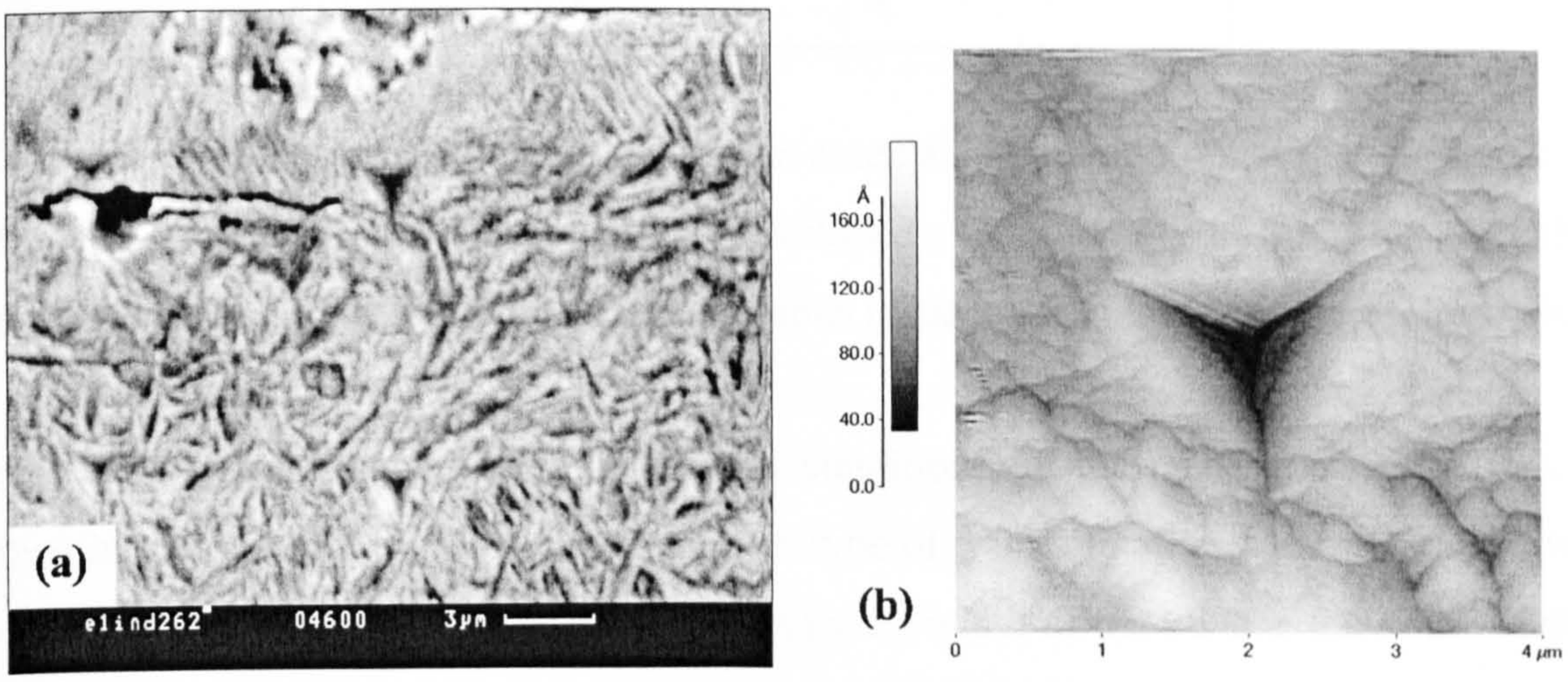


Figure 6-22. (a) SEM image showing an indentation performed in the plastic deformation region, near the crack. (b) AFM image of the indentation.

6.4 Summary

The metallographic investigation of the fatigued specimen has revealed the occurrence of four microstructural features: *dark etching regions* (DER), *white etching bands* (WEB), *butterflies* and, the *plastic deformation layer* (PDL). The hardness values of the different phases found in fatigued specimens are given in Table 6-5.

The dark etching regions initiate at the prior austenite grain boundaries. The DER microstructure is fine and homogeneous as the nanoindentation data indicate. The white etching bands are slightly harder than the original martensite but the space between the

bands is much harder due to intense carbide precipitation in these regions. The plastic deformation layer (PDL) represents work hardened regions near the surface.

Table 6-5. Hardness of different phases present in steel.

Phase	Hardness, H (GPa)
Martensite	10
Dark Etching Regions	12.3
White Etching Bands	11.6
Space between the bands	20.3
Butterflies	12.5
Plastic deformation layer	13.7
Wear debris	6

The butterflies observed in this study are not associated with non-metallic inclusions but they are regions of dark etching phase, developed symmetrically to the prior austenite grain boundaries. The hardness of these regions is identical to the hardness of the DER.

In order to asses the influence of the test conditions on the phase transformations a transformation factor corresponding to each type of transformation has been calculated as the ratio between the width of new phase (p the width of the PDL, d the width of the DER and w the width of the WEB) and the number of stress cycles, N. The values of p, d, and w corresponding to each of the eight experiments are given in Table 6-6.

Table 6-6. The phase transformation factors.

Experiment	$p \cdot 10^{-5}$ (μm/cycle)	$d \cdot 10^{-5}$ (μm/cycle)	$w \cdot 10^{-5}$ (μm/cycle)
1	1.25	51.38	140.28
2	0	19.44	0
3	0.34	4.86	0
4	1.11	5.37	0
5	0	1.11	0
6	2.22	102.78	352.78
7	1.38	38.88	202.78
8	0	3.61	0

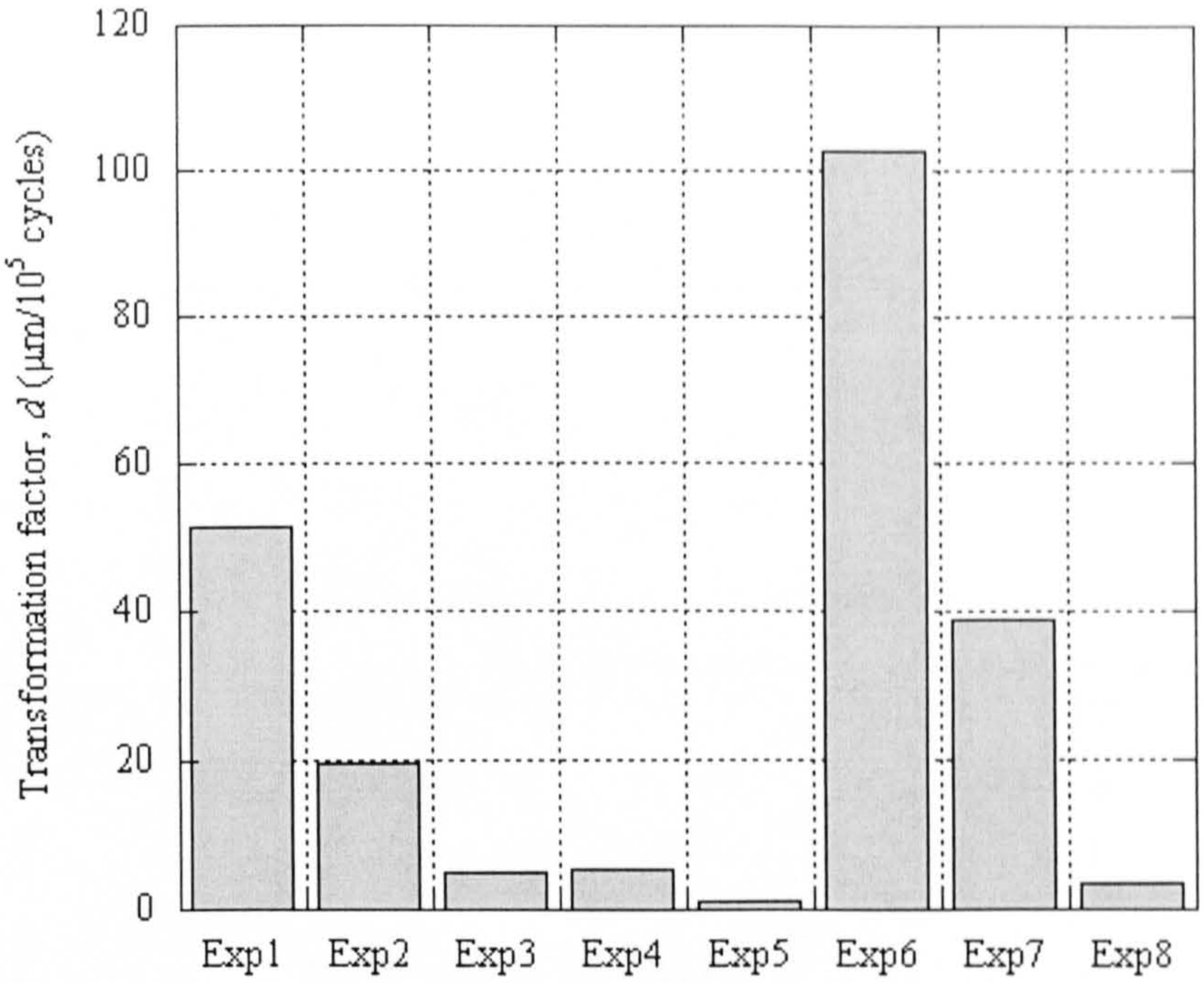


Figure 6-23. The factor of dark etching regions.

The dark etching regions occurred in all tests, therefore the DER factor, *d*, would be more suitable to characterise the phase transformations and the micropitting phenomenon. The columns plotted in Figure 6-23 follows a similar pattern to those for micropitting initiation, from Figure 5-16. This suggests that the micropitting phenomenon and the martensite decay are related to each other.

~ Chapter 7 ~

MICROPITTING IN GEARS

7 MICROPITTING IN GEARS

7.1 General

It has been shown by previous research (Webster & Norbert, 1995) that the twin disc test can be reliably used to simulate the gear tooth contact conditions. In order to be able to compare the micropitting results obtained from the disc tests, which were presented in the previous chapters, similar investigations have been carried out on gears. This chapter presents the results of the investigations carried out on helical gears tested in the Design Unit-Gear Technology Centre from Newcastle University. Emphasis was placed on the metallographic examination of the tested gears.

7.2 Surface topography

Surface topography of the tested gears has been examined by light microscopy and optical profilometry on the xantopren replicas taken from the gear tooth surface after the first testing stage.

The remarks made in section 5.2 regarding the formation of micropits and the surface roughness wavelength in two directions, parallel and perpendicular to the surface lay (or the direction of contact motion) are valid for gears too. Figure 7-1a shows that the first micropits occur on the top of the highest grinding marks and their occurrence follows a certain periodicity, which can be approximated by the surface roughness wavelength in the direction of the surface lay (Ox).

When the micropitting phenomenon is more advanced, as in Figure 7-1b, the occurrence of micropits follows a certain periodicity in the direction perpendicular to the surface lay (Oy) which can be determined by the surface profile wavelength in this given direction.

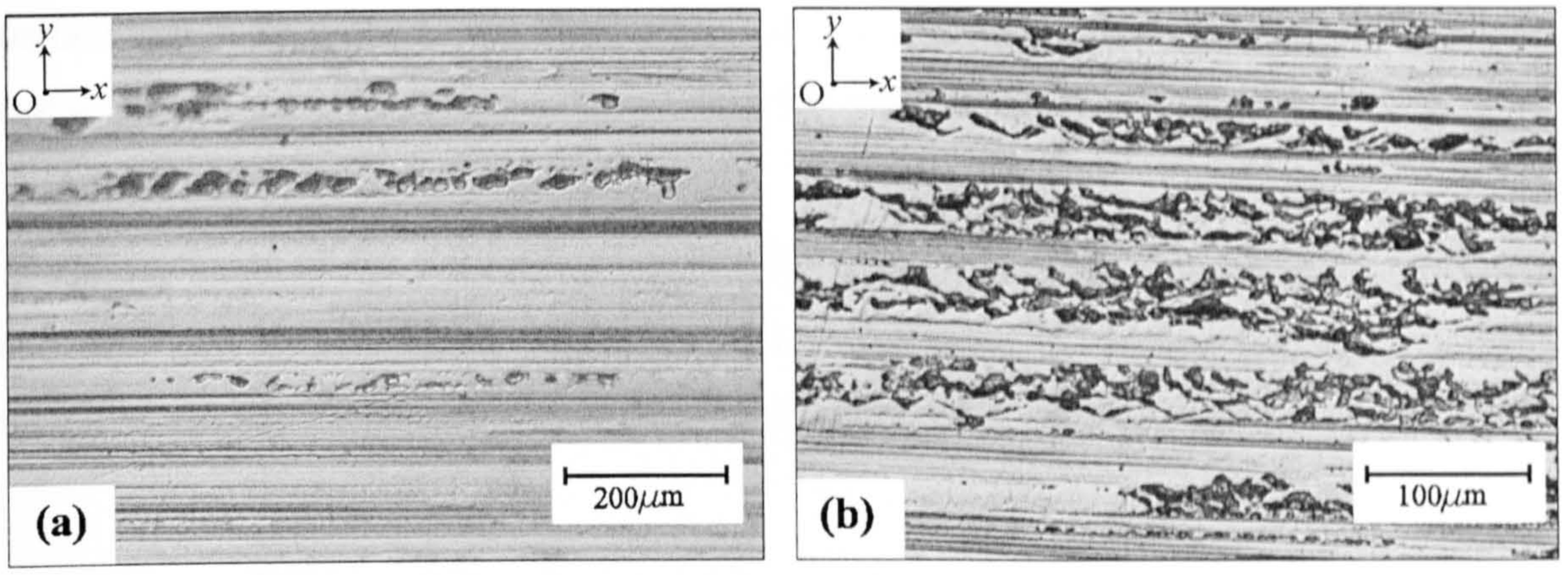


Figure 7-1. Light microscopy on replicas. Micropitting initiation. (a) Gear C15, addendum region on the wheel, $5 \cdot 10^6$ cycles. (b) Gear B29, dedendum region on the wheel, $5.12 \cdot 10^6$ cycles.

Similar to the micropits observed on the disc specimens, the micropits observed on gear tooth surface have a shape elongated in the direction of the surface lay (see Figure 7-2a & b). It has been shown in the previous chapter that this is due the shape of asperities, which have an elliptical geometry.

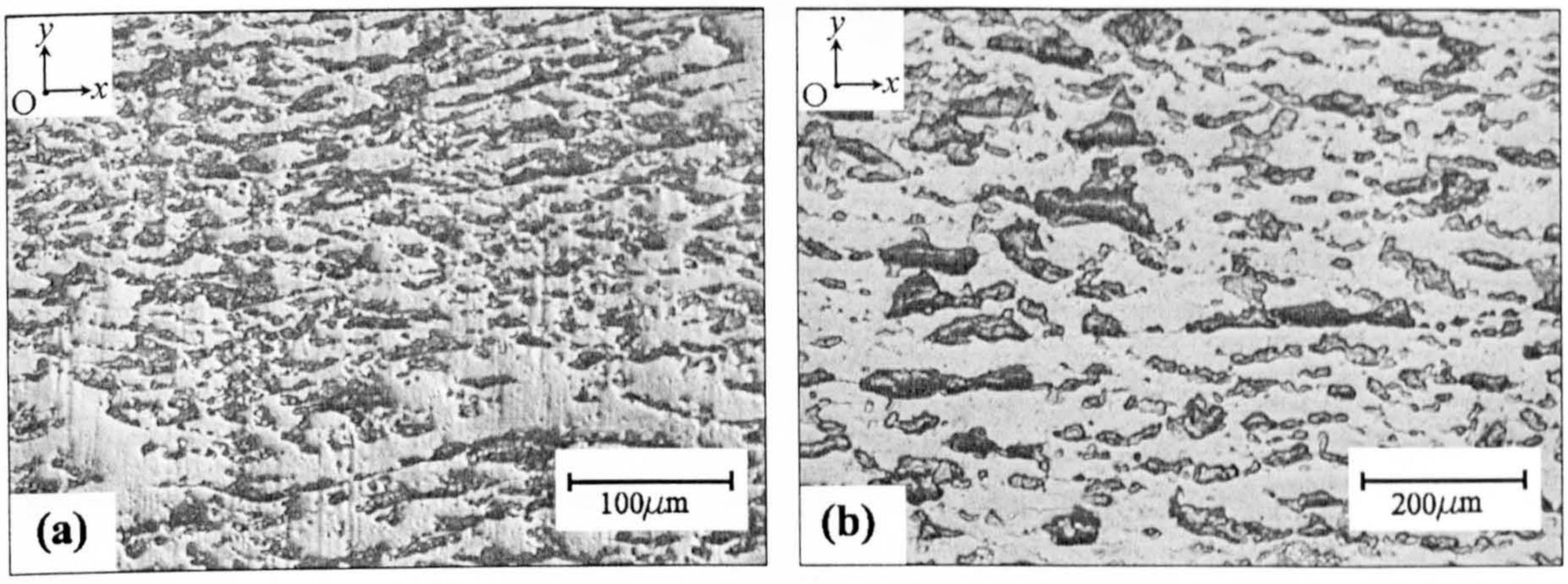


Figure 7-2. Light microscopy on replicas. Gear B12, addendum region on the wheel, $5.12 \cdot 10^6$ cycles.

Because the replicas represent the mirror image of the actual surface of the gear tooth, the micropits are seen as protuberances and not as cavities. The advantage is that the shape of the micropits can be better visualised. The optical profilometer image from Figure 7-3a shows an isolated micropit observed on replica taken from gear B29 after the first testing stage ($5.12 \cdot 10^6$ cycles). The initial microcrack was initiated in A and after a growth at an inclined angle, from A to B, the crack propagated parallel to the

surface (BC). In point C the crack changed the direction of propagation towards the surface and when it had reached the surface the micropit has been formed.

The dimensions of the micropit are given in Figure 7-3b & c; they are 28.4 μm width and 40.7 μm length. These values provide information about the shape of asperity, which as already stated, is elongated in the direction of the surface lay.

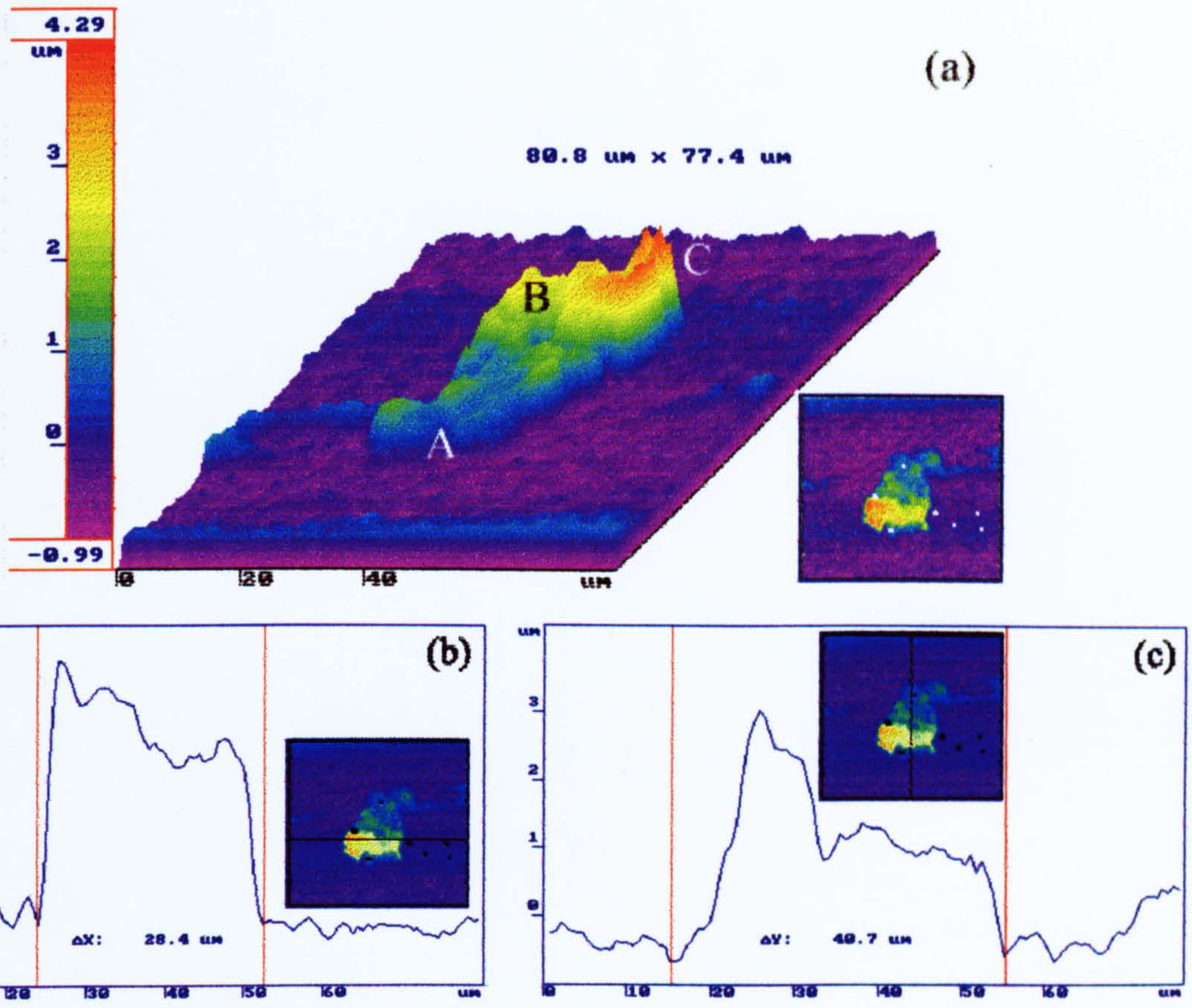


Figure 7-3. Optical profilometer images. (a) Three dimensional view of the micropit. (b) and (c) Surface profiles in two orthogonal directions. Gear B29.

During the initiation period the size of the microcracks might be very small, only a few nanometers. Micropits in the initiation stage can be seen in Figure 7-4. The space between the surface valleys is filled with a large number of very small micropits. Each micropit can be associated with an asperity found on the gear tooth surface.

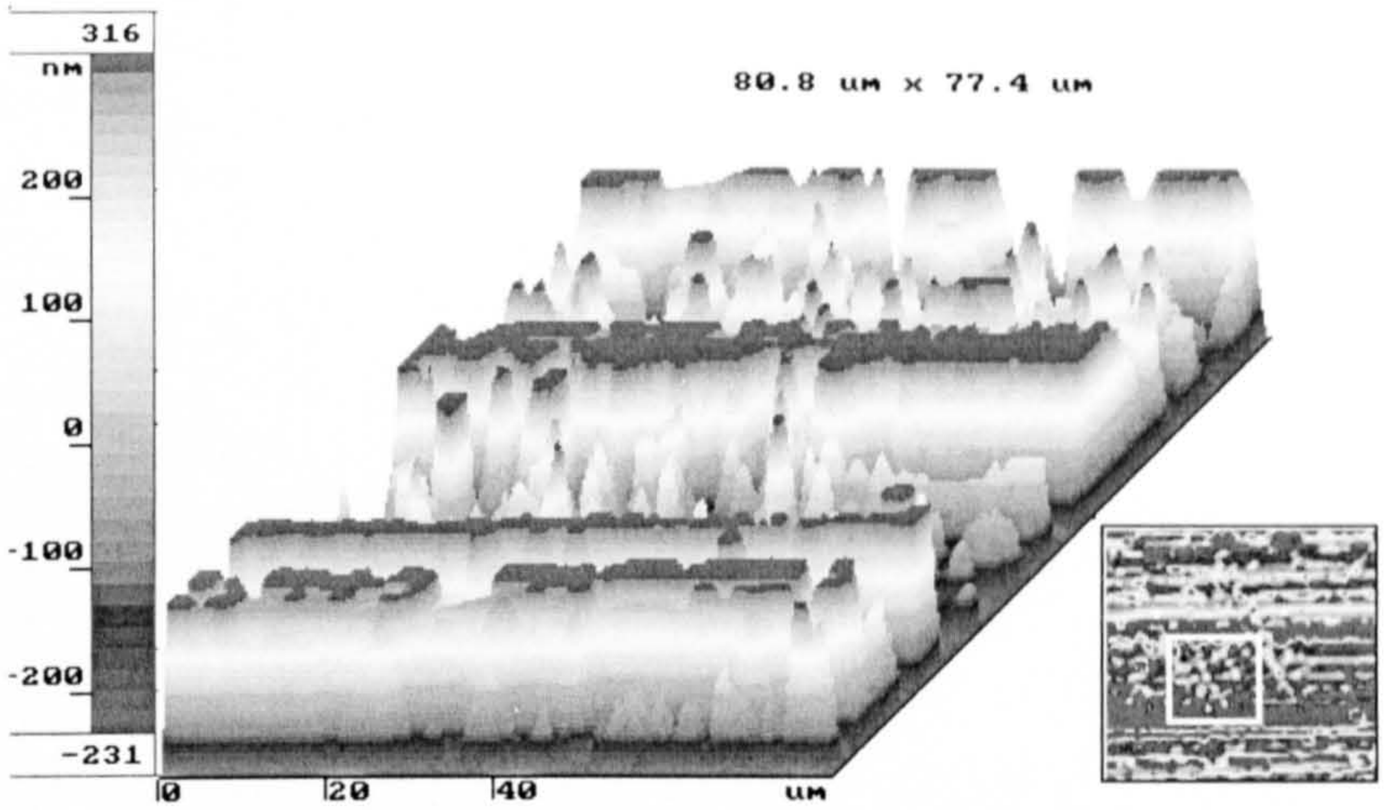


Figure 7-4. Optical profilometer image showing micropits in the initiation period. Gear C76, dedendum region.

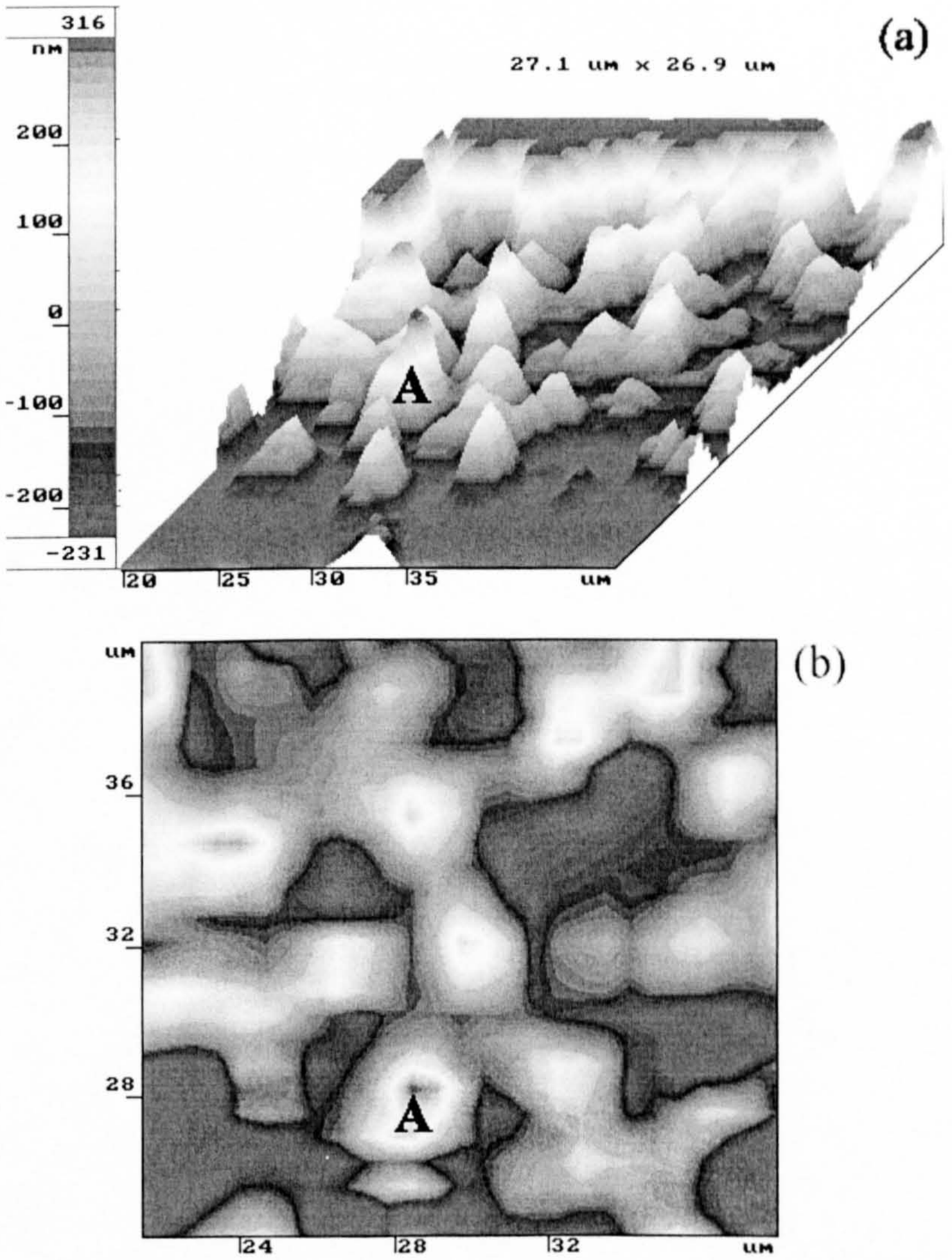


Figure 7-5. Optical profilometer image. (a) Three dimensional view of the region framed in Figure 7-4. (b) Top view of the same region.

The region inside the white square in Figure 7-4 is further magnified in Figure 7-5a & b. The micropits are linked together by bridges, which represent very small micropits and they probably grow during cycling giving rise to bigger and deeper micropits.

The mechanism of micropitting propagation proposed in Chapter 5, which applies to discs, can also be applied to gears. As cycling continues, the micropits grow and merge with their neighbours resulting in bigger and deeper micropits (see Figure 7-6). At the same time, the grinding marks tend to flatten and new micropits form on the highest marks.

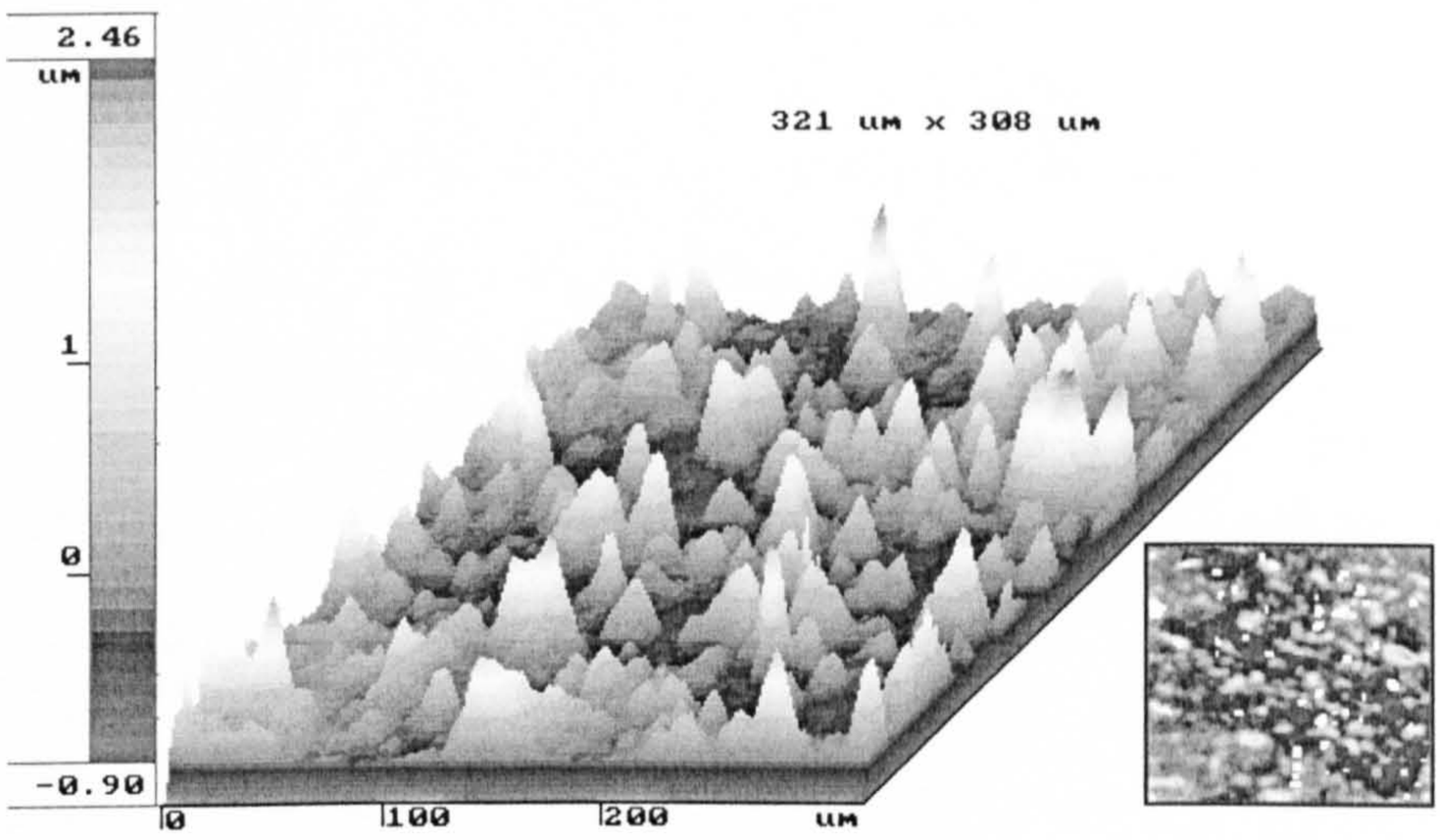


Figure 7-6. Optical profilometer image showing micropits on the tooth surface. Gear B34, addendum region.

7.3 Initiation and propagation of cracks

Compared to the cracks observed on the disc specimens, the microscopic investigation of the polished cross sections taken from the gear tooth has revealed similar aspects of the surface initiated cracks. The main difference is that the subsurface initiated cracks observed on discs, which propagate parallel to the surface have not been observed in gears.

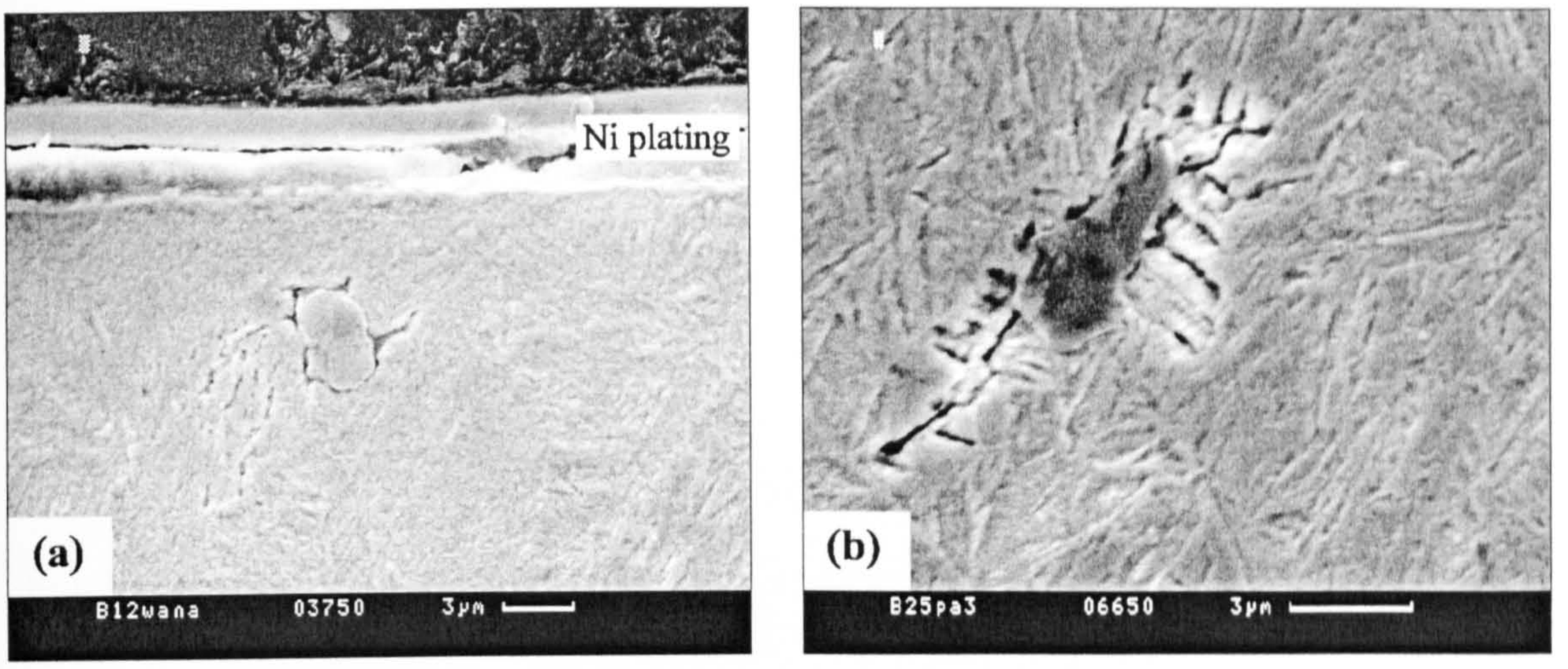


Figure 7-7. Cracks initiated at non-metallic inclusions. (a) Gear B12. (b) Gear B25.

Moreover, the few cracks initiated in the subsurface region are always associated with non-metallic inclusions (see Figure 7-7) but none of the cracks observed has reached the surface. As for the micropitting in discs, it can be concluded that micropitting in gears is a phenomenon driven by the propagation of surface initiated cracks.

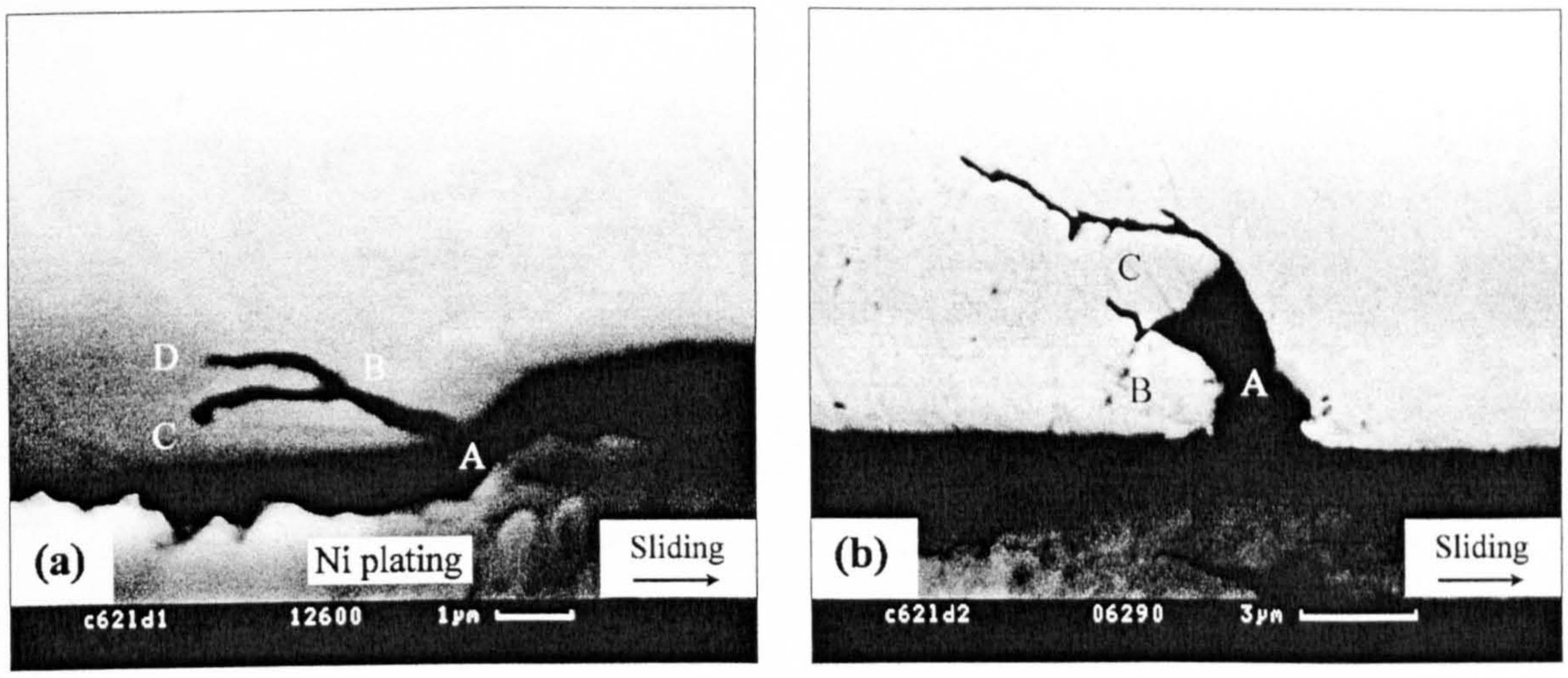


Figure 7-8. SEM images showing microcracks in gear C62.

The mechanism of crack propagation is similar to that described in section 5.7. It can be seen in Figure 7-8a that the cracks initiate at the edge of asperities (point A). The crack propagates at a shallow angle to the surface in the direction opposite to that of sliding and in B one branch of the crack starts to propagate parallel to the surface (BC) and the initial crack continues to propagate into the bulk (BD).

The mechanism of crack propagation explains the coalescence of micropits. The micropit marked A in Figure 7-8b will increase its volume by the removal of the volumes marked as B and C.

Another eloquent example is given in Figure 7-9a. The size of the micropit marked with A corresponds to the width of the grinding marks. The region beneath was exposed to high contact stresses and the initiation of a crack became possible. During the propagation of the initial crack, which generated the pit A, branching has occurred and two branches (BC and DE) propagated into the bulk. These new cracks serve as sources for other branches (FG, HI and MN), which propagate towards the surface, contributing in this way to the micropit coalescence.

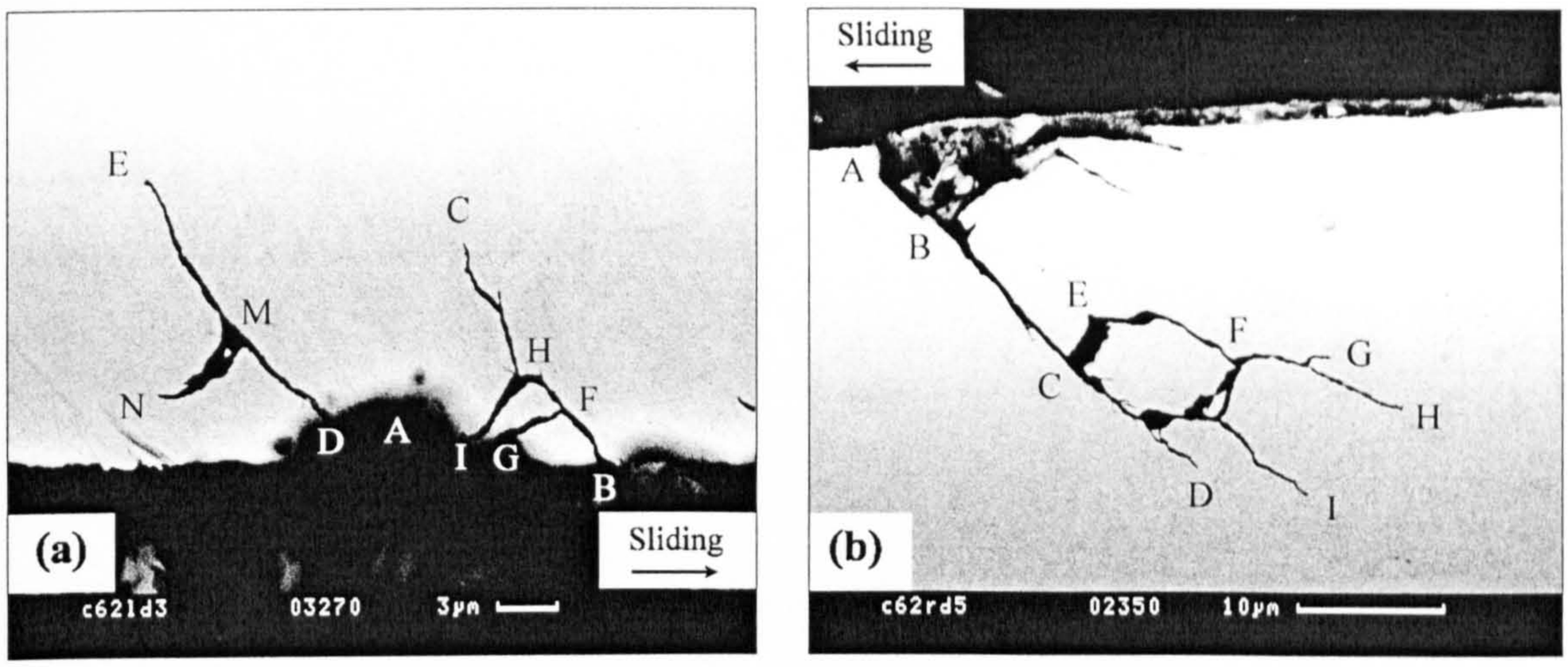


Figure 7-9. SEM images showing microcracks in gear C62. (a) Crack branching occurs at the asperity extremities. (b) Crack propagate at the prior austenite grain boundary.

Figure 7-8b illustrates the case of micropitting leading to macropitting. The original crack was initiated at the surface, in point A, after a growth of few microns a new crack has been formed in point B, which caused the formation of a micropit. The initial crack continued to propagate at the same angle and, in point C, branching occurred. Judging by the circular shape of the path followed by the crack, it is likely that the crack propagated at the prior austenite grain boundary. It has been shown in Chapter 6 that DER initiate at the prior austenite grain boundary. The crack path from Figure 7-8b resembles to that from Figure 5-23b, which was obtained from a disc specimen. Both, the initiation of DER and the propagation of cracks at the prior austenite grain boundary suggest that this region is the most critical.

7.4 Phase transformations

The investigations carried out on the disc specimen surface and gear tooth surface, have revealed similarities concerning the size, shape and distribution of the micropits. On the other hand, the study of the associated microcracks suggests similar mechanisms of crack initiation and propagation both, in discs and gears. The metallographic investigation of the micropitted discs have revealed that the micropitting phenomenon is associated with a number of phase transformations, some of them (DER, WEB, and butterflies) also reported to occur in fatigued bearings.
The question that arises is whether these microstructural changes occur in gears.

7.4.1 Dark etching regions

The metallographic examination has revealed the same etching characteristics as in disc specimens. The dark etching regions (DER) as a continuous band underneath the gear tooth surface, have been observed only in gear C44 in the dedendum regions (see Figure 7-10a). Most often the DER phase is seen as isolated zones beneath the surface, as shown in Figure 7-10b.

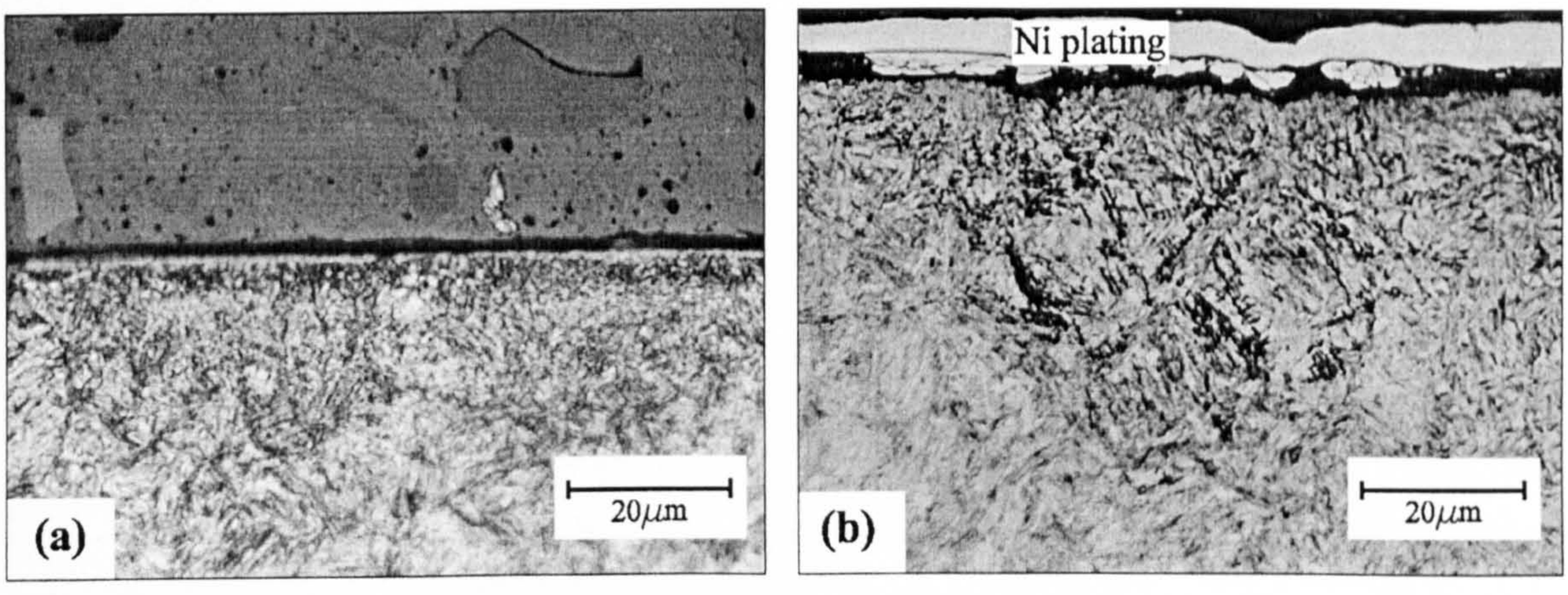


Figure 7-10. Light microscope images showing dark etching regions. (a) Gear C44, dedendum. (b) Gear B25 addendum.

The idea of DER initiating at the prior austenite grain boundary is supported by the micrographs from Figure 7-11a &b in which the prior austenite grain has been marked with the letter A.

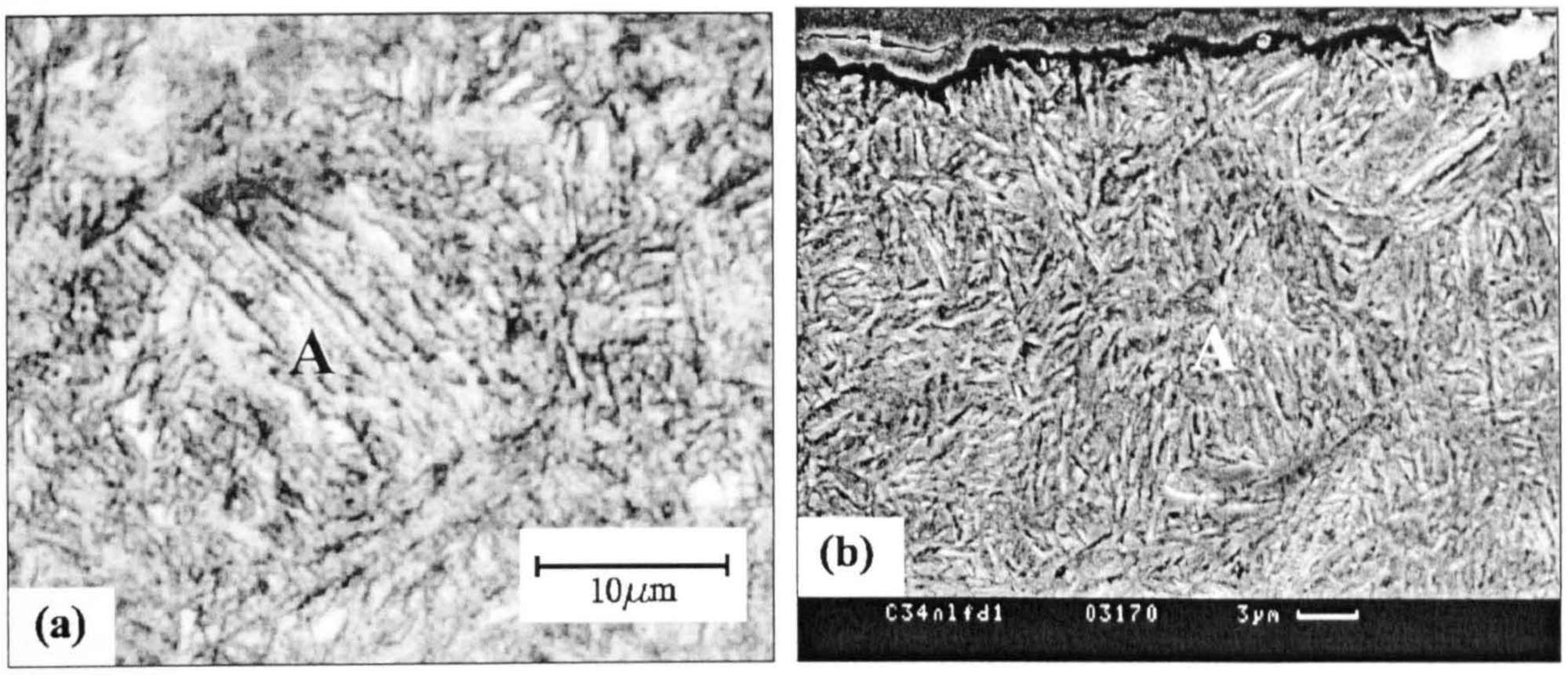


Figure 7-11. DER nucleation at the prior austenite grain boundary. (a) Light microscope image. (b) SEM image. Gear C34. Nital etch.

The newly nucleated grains grow and break down the martensite needles until the entire martensitic structure, within the original austenite grain, is transformed, as shown in Figure 7-12a &b. The biggest grains appear arranged in a circle which suggest their development has started from the boundary of the original austenite.

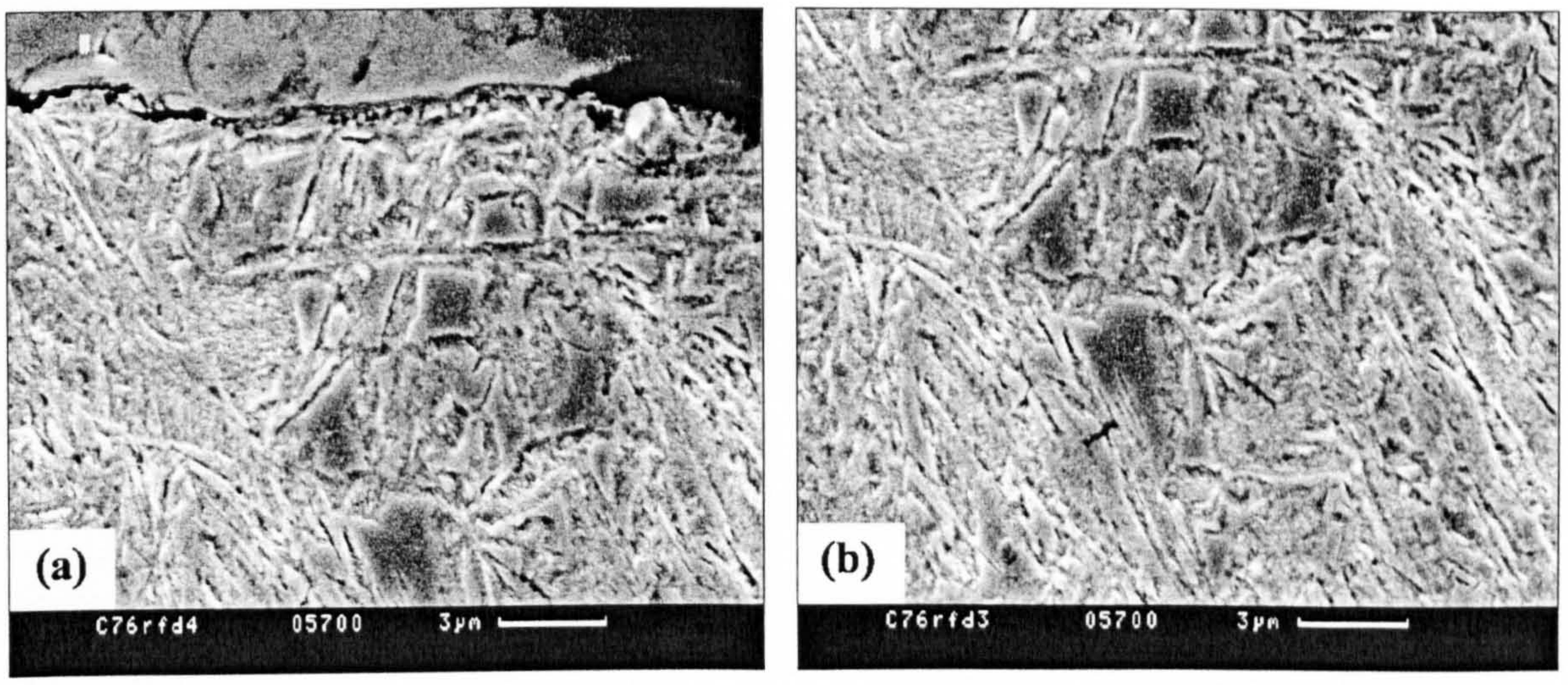


Figure 7-12. Grains of dark etching phase. Gear C76. Nital etch.

Figure 7-13 shows the boundary region between the dark etching phase and the non-transformed martensite. The microstructure resembles to that shown in Figure 6-5b. The presence of carbide precipitates, in Figure 7-13 and 6-5b, indicates that the mechanism of DER formation is similar in gears and discs. The precipitation takes place by carbon

diffusion and this implies that the thermal effects should have a great importance in the development of DER.

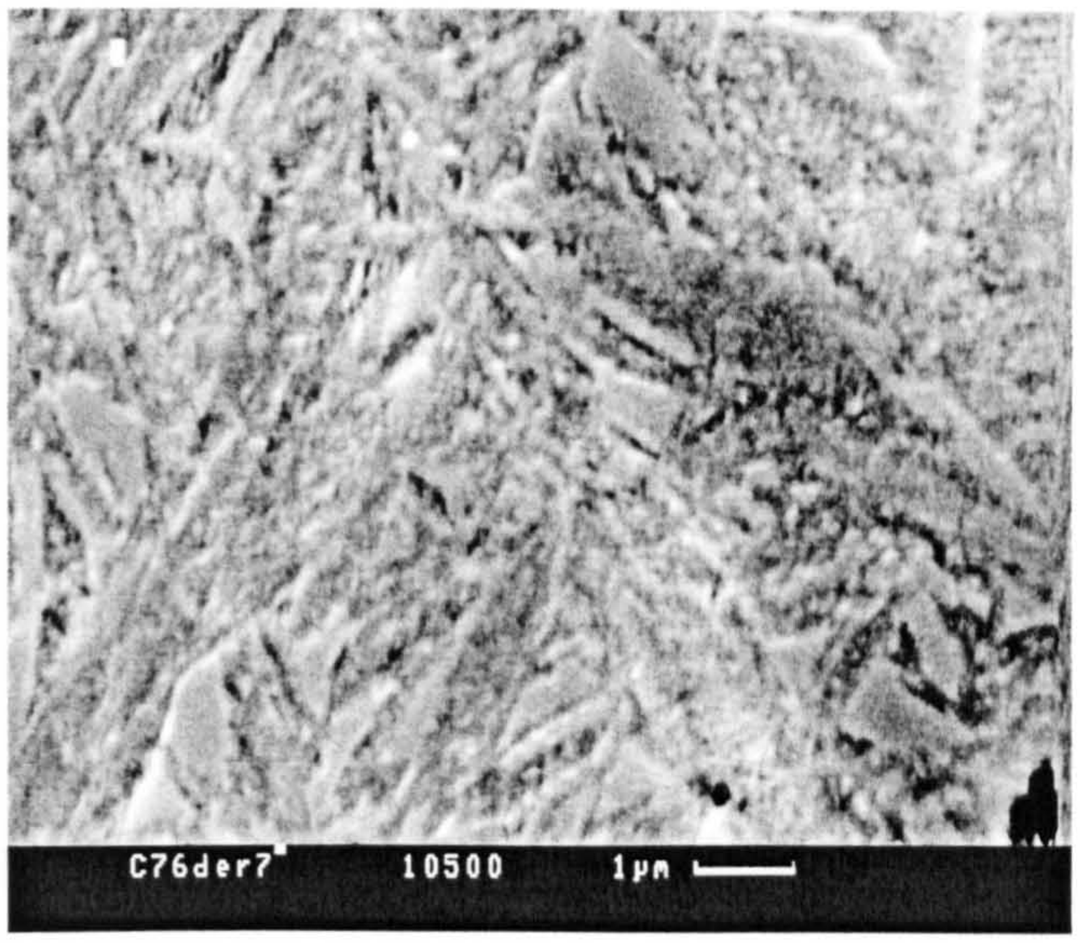


Figure 7-13. The boundary region between DER and non-transformed material. Gear C76; Nital etch.

7.4.2 White etching bands

Although a texture well developed as in discs (see the previous chapter) has not been observed in gears, a certain directionality of the microstructure has been noted in almost all specimens. Two examples are shown in Figure 7-14 and 7-15. The bands from the pictures shown below resemble with those from Figure 6-7, which suggests they are white etching bands in the early stage of development.

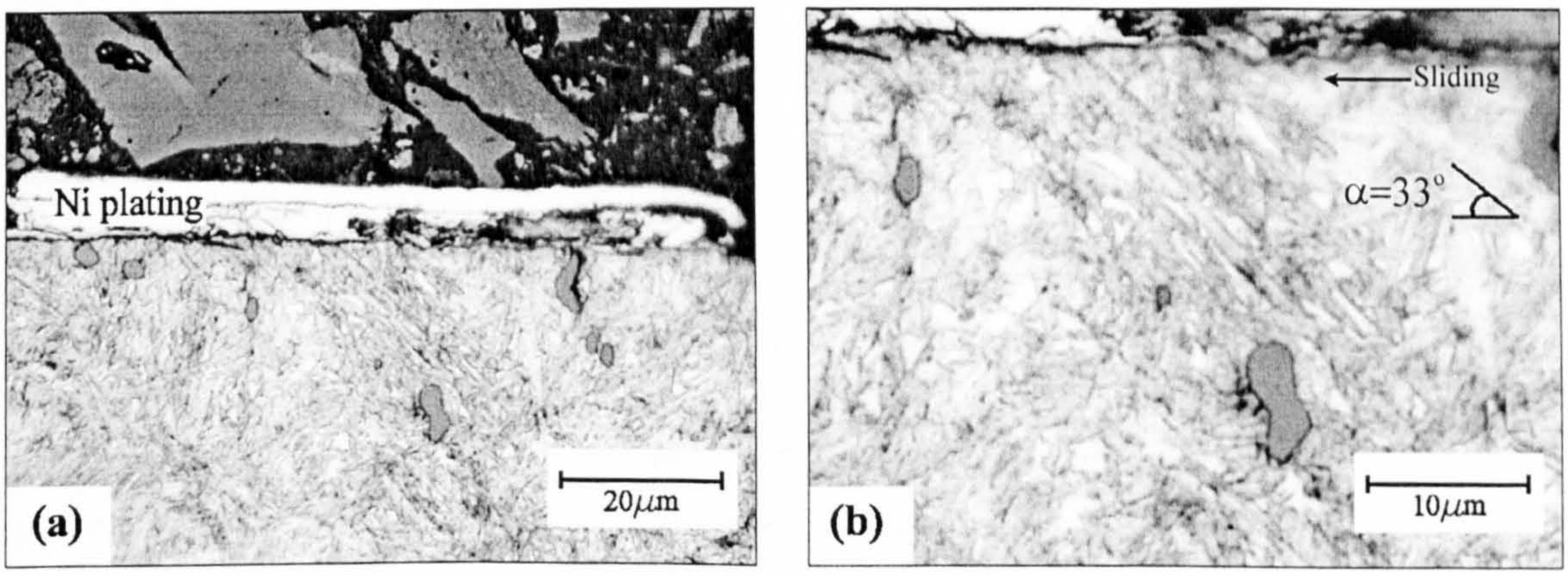


Figure 7-14. Light microscope images. (a) Near surface region in gear B29. (b) Higher magnification of the WEB region. Nital etch.

The angle of inclination to the surface $\alpha = 33°$ in Figure 7-14 and $\alpha=42°$ in Figure 7-15. The value of the angle is close to the angle of crack propagation.

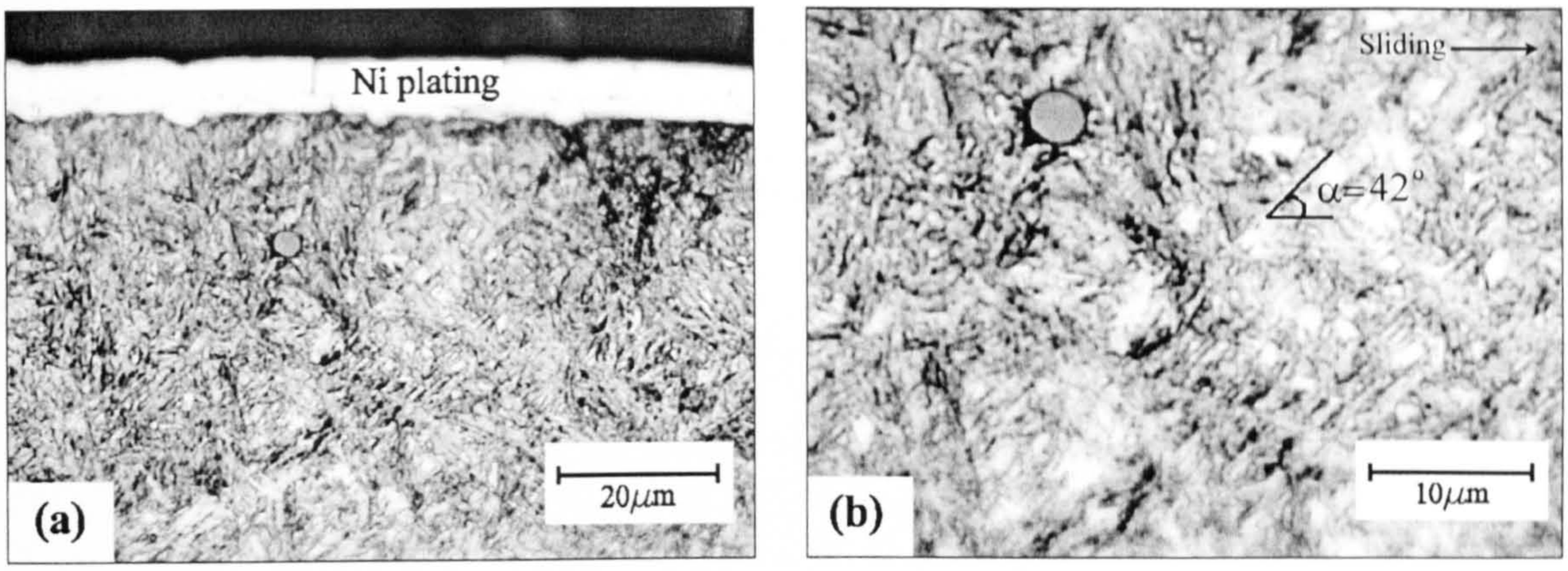


Figure 7-15. Light microscope images. (a) Near surface region in gear C34. (b) Higher magnification of the WEB region. Nital etch.

As mentioned in Chapter 6, the angle of the white etching bands depends on the sliding speed and direction, which in gears, have opposite directions in dedendum and addendum and vary from zero at the pitch line.

7.4.3 Plastic deformation layer

The common feature of all gears on which micropitting has been observed, is represented by small semi-circular regions with a diameter in the order of tens of microns, in the proximity of the surface, which exhibit a very fine non-martensitic microstructure. Two examples are given in Figure 7-16.

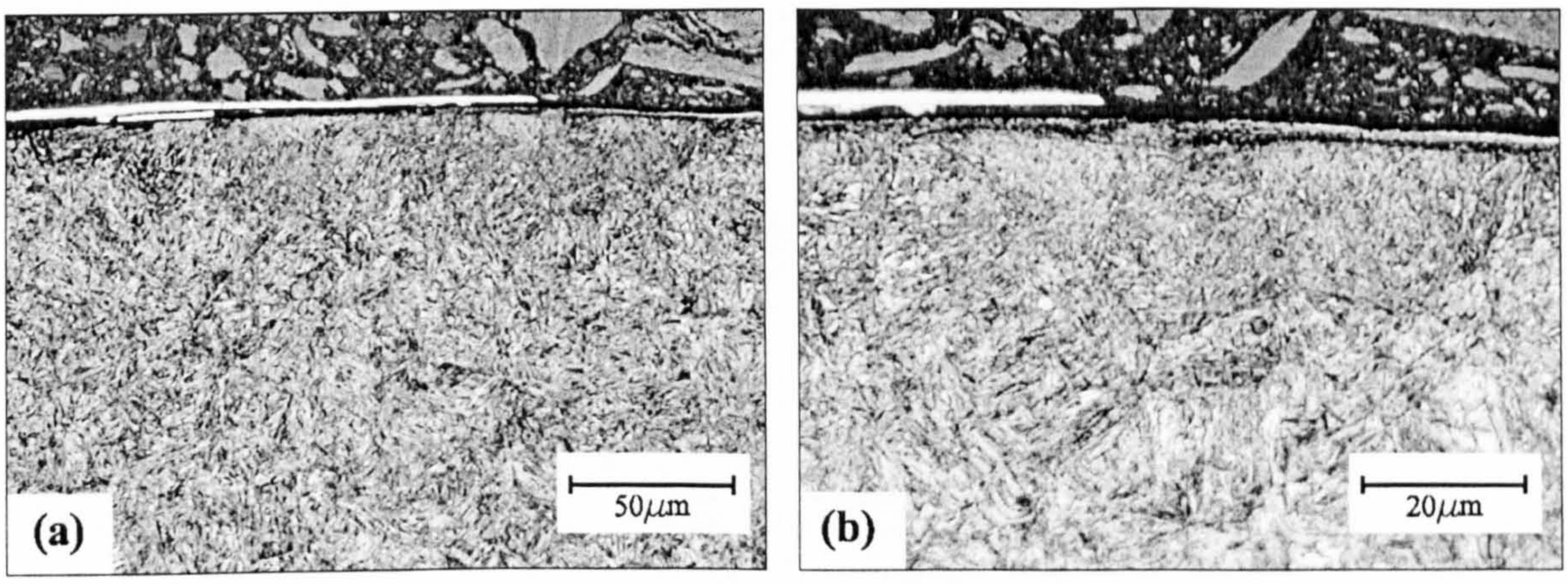


Figure 7-16. Light microscope image of the near surface region showing the plastic deformation zone. (b) Higher magnification. Gear B12, addendum. Nital etch.

The picture from Figure 7-16b is a higher magnification of the region close to the surface in Figure 7-16a. This pattern is similar to the plastic deformation layer (PDL) observed in the disc samples.

Figures 7-17 a & b depict another plastic deformation region found in the addendum region of the gear B25. The right side of micrograph (a) is further magnified in (b). It can also be seen that the dark etching regions give the contour of the plastic deformation region. A surface initiated crack propagates along the boundary region. The conclusion that can be drawn is that the boundary of these severely deformed regions represents a critical zone. The DER phase is seen to initiate in this zone and also the cracks. Similar aspects were observed in the disc specimens (see Figure 6-19). It can be concluded that the mechanism of micropitting in gears is similar to that in discs. A micropit forms when a crack initiated at the surface, at the point of intersection between the surface and the boundary of the plastic deformation region, it then propagates along the boundary, until it reaches the surface again.

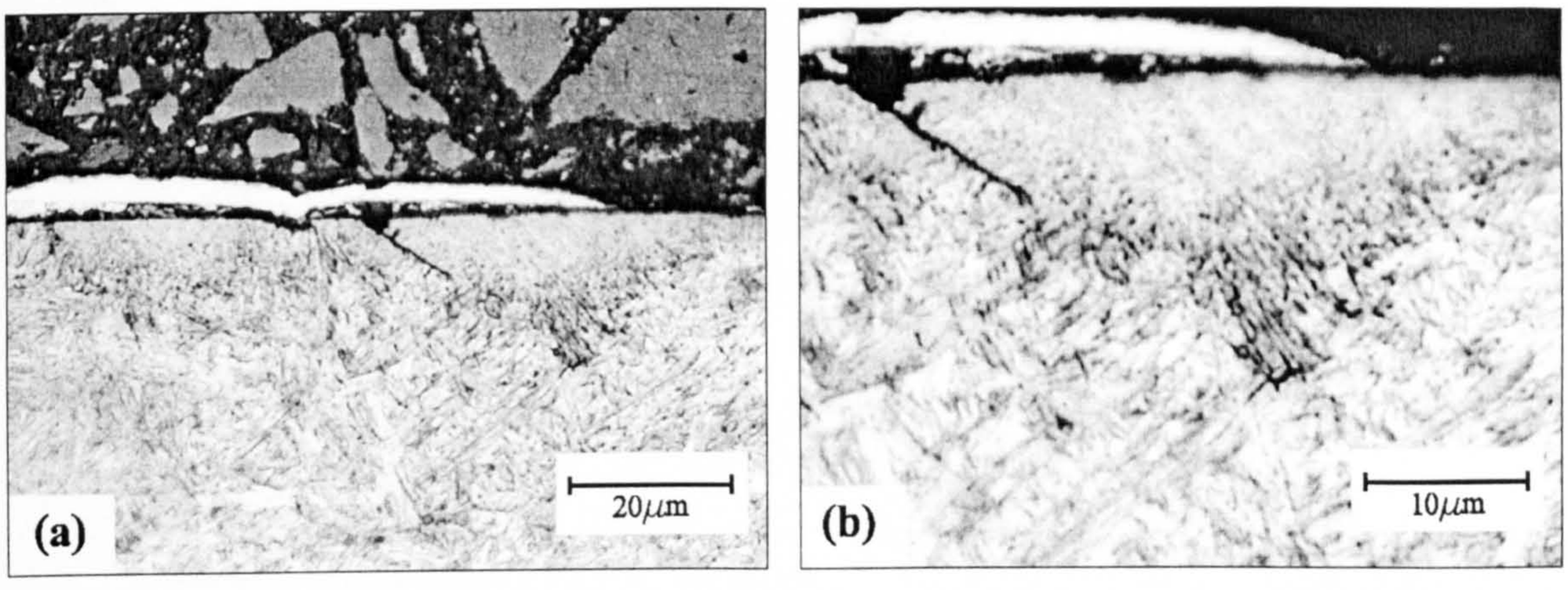


Figure 7-17. Light microscope image of the near surface region showing the plastic deformation zone bordered by dark etching regions. (b) Higher magnification. Gear B12, addendum. Nital etch.

The mechanical properties of the plastic deformation region and its boundary have been determined by nanoindentation. Two tests have been performed in the regions situated to the left and the right side of the crack from Figure 7-17a after re-polishing the sample. The arrays used in the nanoindentation test are shown in Figure 7-18a & b. The values of Young's modulus and hardness obtained from the array shown in Figure 7-18a are given in Table 7-1.

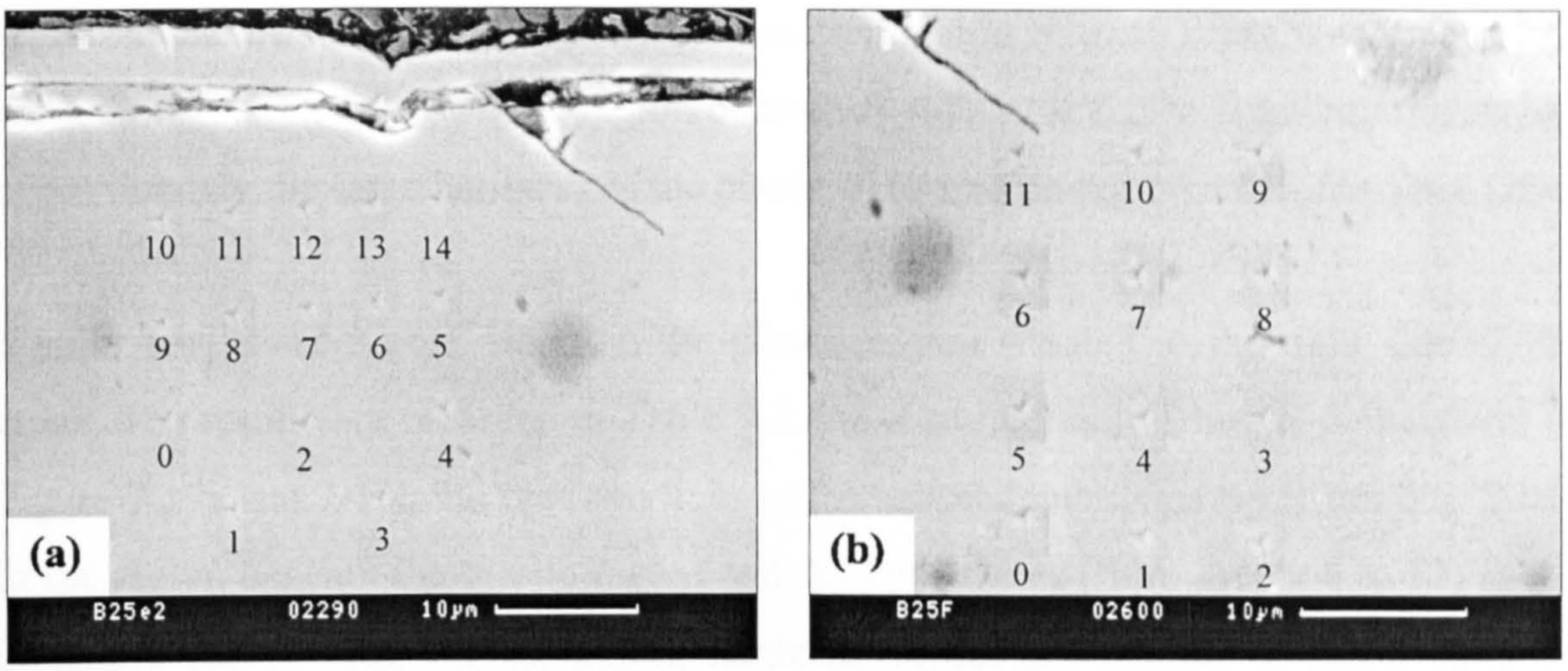


Figure 7-18. Nanoindentation arrays used to determine the mechanical properties of the plastic deformation region and its boundary.

Table 7-1. Nanoindentation data corresponding to the indentations shown in Figure 7-18a.

Indentation	Young's modulus E (GPa)	Hardness H (GPa)
1	166.5	19.9
2	156.4	18.4
3	167.8	16.9
4	188.3	16.3
5	188.9	15.5
6	204.1	14.1
7	203.5	14.3
8	200.0	13.3
9	208.4	13.7
10	210.8	13.9
11	212.0	13.9
12	214.6	14.8
13	211.3	13.8
14	201.4	12.7

The first line of indents (from 0 to 4) is located beneath the plastic region, the second line (from 5 to 9) at the boundary and the third line within the plastic deformation

region. It can be seen that the plastic deformation region is much softer with an average hardness H = 13.8 GPa than the region beneath H = 17.8 GPa. The boundary region has approximately the same hardness as the plastic deformation region itself, H = 14.1 GPa.

Similar results were obtained from the plastic region situated to the right side of the crack. The results are recorded in Table 7-2. As it can be seen from the comparison of Figure 7-17a and 7-18b the first two lines of the nanoindentation array (from 0 to 5) are in the region beneath the plastic region and the upper lines (from indent 6 to 11) in the boundary region. The corresponding hardness values are similar to those obtained previously, i.e., H = 17.2 GPa for the region beneath and H = 13.7 GPa for the plastic deformation region. These results are in agreement with those obtained for the disc specimens. A micropit forms due to the crack propagation at the boundary between a softer phase - the plastic deformation region (PDR), H = 13.7 GPa and the harder phase situated beneath, H = 17.5 GPa. This statement is supported by the nanoindentation data given in Table 7-3, obtained form the indents shown in Figure 7-19a.

Table 7-2. Nanoindentation data corresponding to the indentations shown in Figure 7-18b.

Indentation	Young's modulus E (GPa)	Hardness H (GPa)
1	165.5	17.7
2	167.6	19.7
3	172.1	17.8
4	184.2	15.2
5	188.4	15.8
6	194.0	13.8
7	197.8	13.6
8	206.3	12.9
9	197.5	14.3
10	208.7	13.5
11	211.5	14.1

The crack propagates between a hard region, to the left (indentation 2, H = 18.4 GPa) and a soft one, to the right (indentation 3, H = 14.2 GPa). The indentation performed near the crack tip (indent 2) is shown in Figure 7-20.

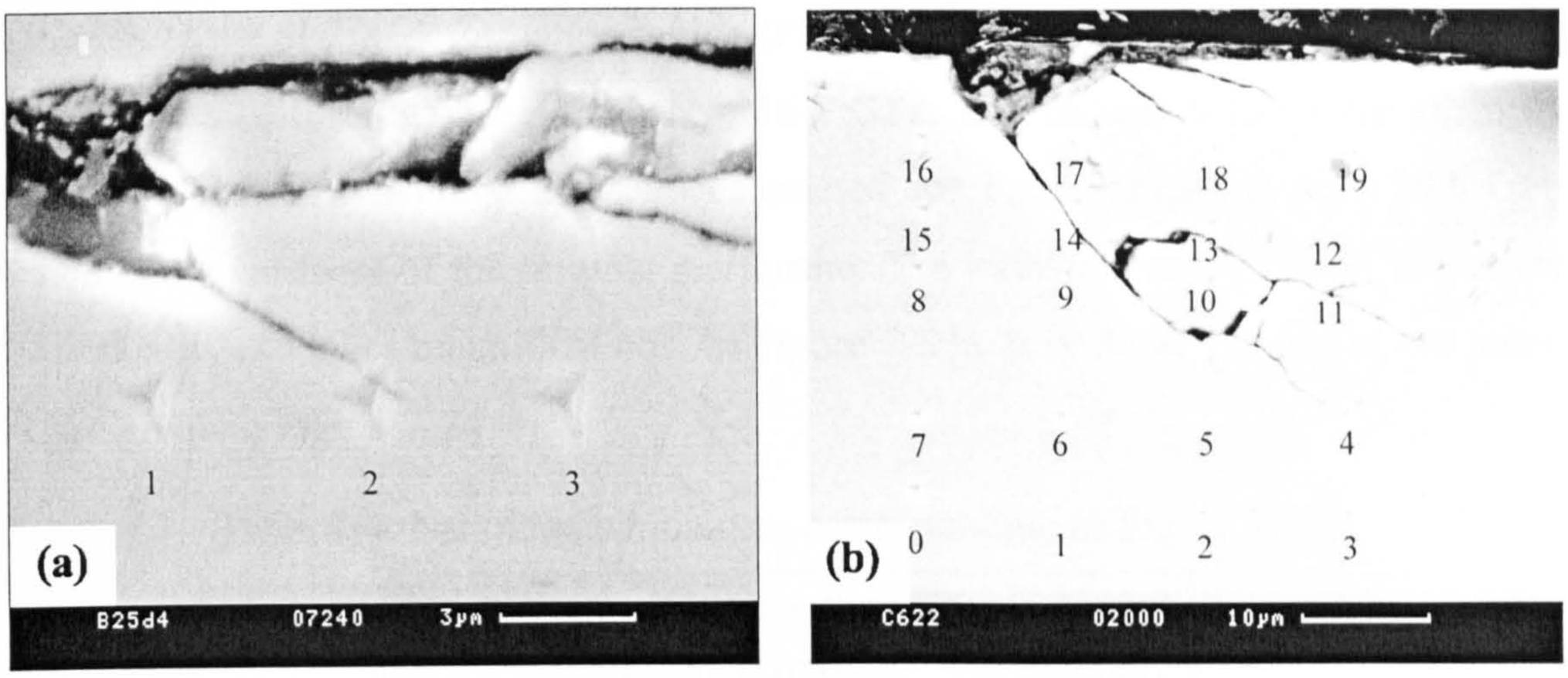


Figure 7-19. SEM images of the nanoindentation arrays used to determine the mechanical properties in the near surface regions. (a) Gear B25. (b) Gear C62.

Table 7-3. Nanoindentation data corresponding to the indentations shown in Figure 7-19a.

Indentation	Young's modulus E (GPa)	Hardness H (GPa)
1	170.6	14.2
2	155.5	18.4
3	150.7	18.2

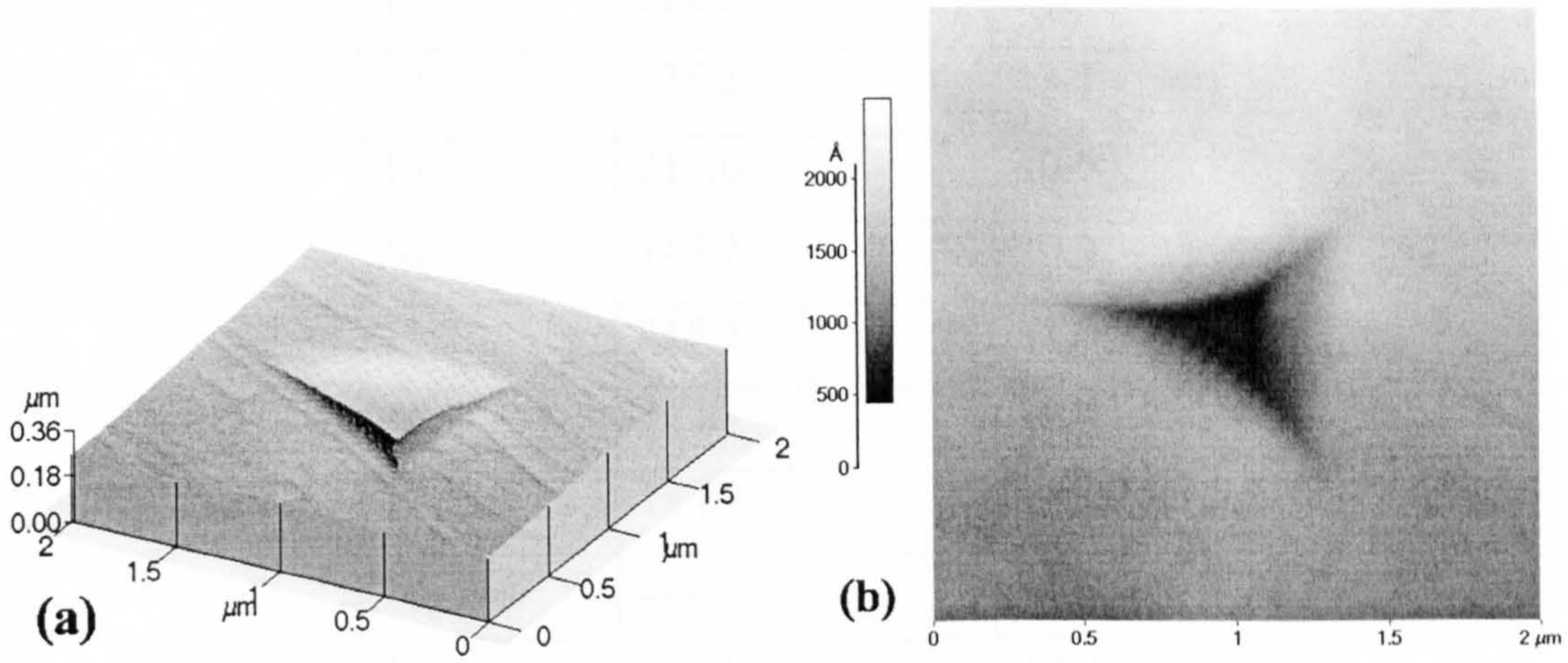


Figure 7-20. AFM image of the indentation no. 2 from Figure 7-19a.

The results of the nanoindentation test performed according to the array shown in Figure 7-19b, which are given in Table 7-4, indicate that the boundary between the plastic deformation region (PDR) and the region below is softer than both phases. The hardness of the plastic deformation region (PDR) is H = 12.7 GPa (indents 16 to 19) and the hardness of the region below is H = 16.2 GPa. The indents 9 to 14 are all in the critical region. The average hardness calculated for these indents is H = 10.6 GPa, which is the hardness of the original martensite. The indentations 10 and 11 are shown individually, at higher magnifications, in Figure 7-23a & b. AFM images of the indent 11 are given in Figure 7-24.

Table 7-4. Nanoindentation data corresponding to Figure 7-19b.

Indentation	Young's modulus E (GPa)	Hardness H (GPa)
1	152.3	29.2
2	156.6	26.1
3	171.0	21.6
4	175.9	17.8
5	191.7	16.7
6	203.4	15.8
7	210.1	14.7
8	261.9	12.2
9	258.6	11.0
10	219.1	11.9
11	212.3	9.4
12	235.3	13.1
13	215.6	10.5
14	145.2	10.2
15	215.2	11.0
16	215.3	12.4
17	184.2	12.4
18	223.5	13.1
19	230.9	12.9

It is likely that a crack will propagate within the soft boundary region, although cracks propagating within the plastic deformation region or within the harder region have been observed. Figure 7-21 shows an SEM image of the indentation 4 in (a) and an AFM image in (b). The hardness near the crack tip, H = 17.8 GPa is similar to other values obtained at the same depth.

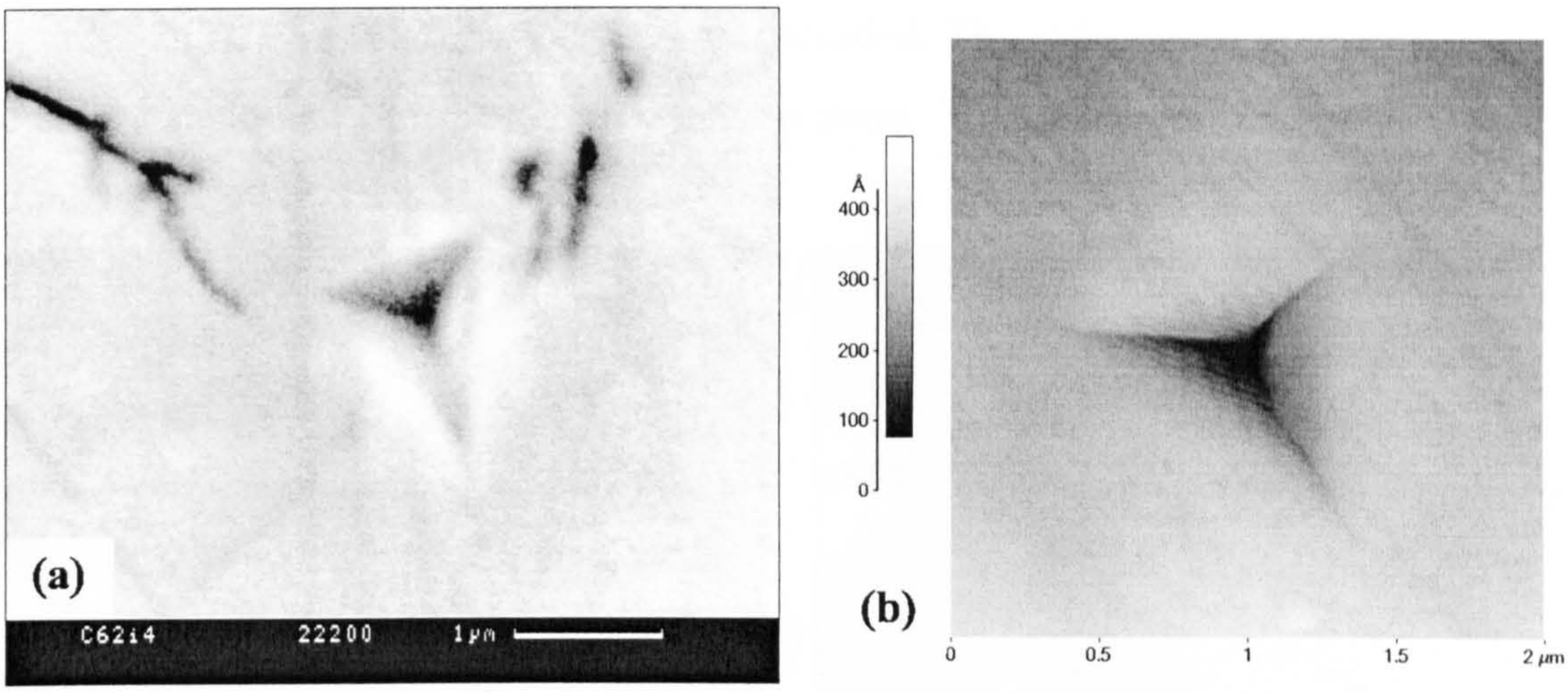


Figure 7-21. Indentation 4. (a) SEM image. (b) AFM image.

Indentation 18, which is shown in Figure 7-22 (SEM image in (a) and AFM image in (b)) has been performed near the tip of a crack propagating within the plastic deformation region. The determined hardness is in the range of the values obtained for the plastic deformation region.

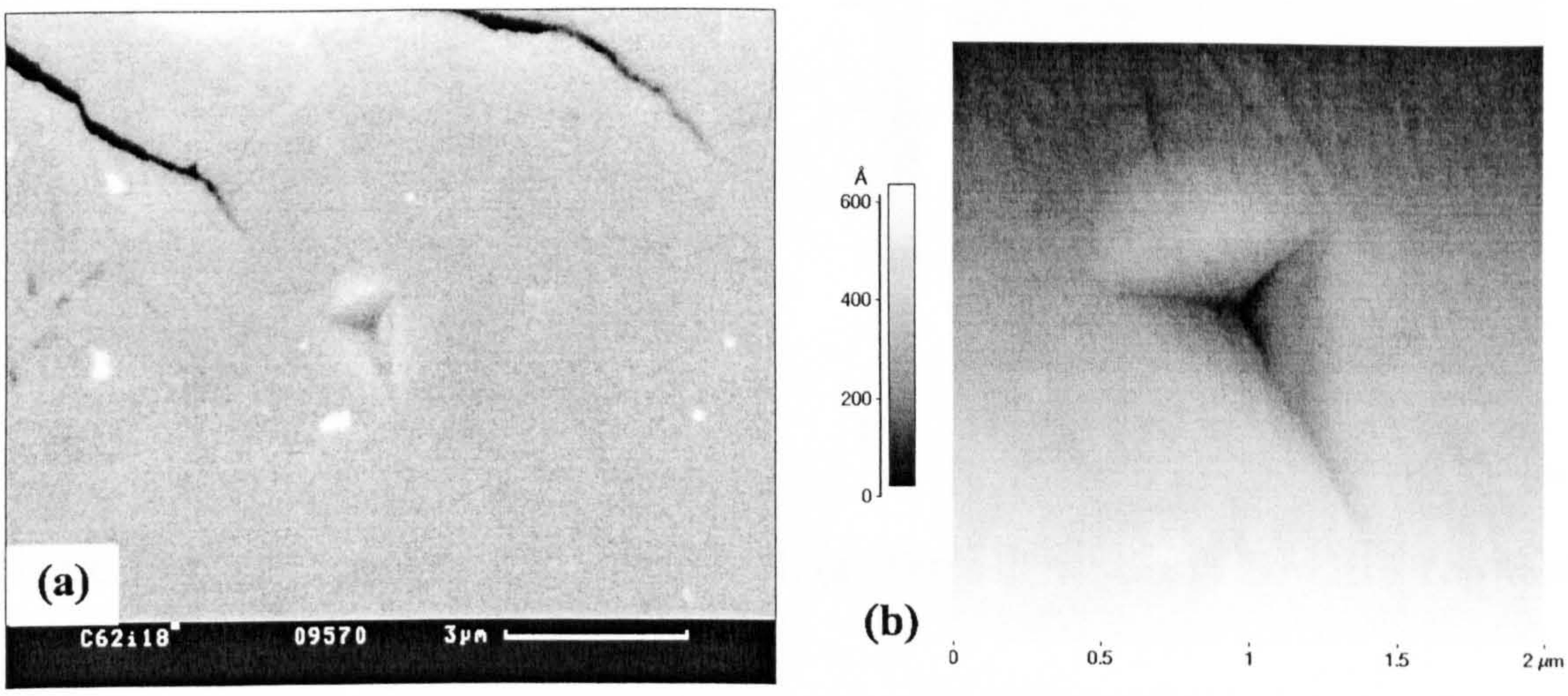


Figure 7-22. Indentation 18. (a) SEM image. (b) AFM image.

The indentation number 10 (see Figure 7-23a) has been performed in centre of the circle described by the path followed by the crack. The hardness value, H = 11.9 GPa corresponds to the hardness of DER. This suggests that the crack propagated along the boundary of a prior austenite grain, within which the martensite has been transformed into dark etching phase. The hardness determined from indentation number 11 (see Figure 7-23b), H = 9.4 GPa is the lowest recorded. The indentation 11 is located in the boundary region of the plastic deformation zone. This confirms, once more, that the boundary is softer.

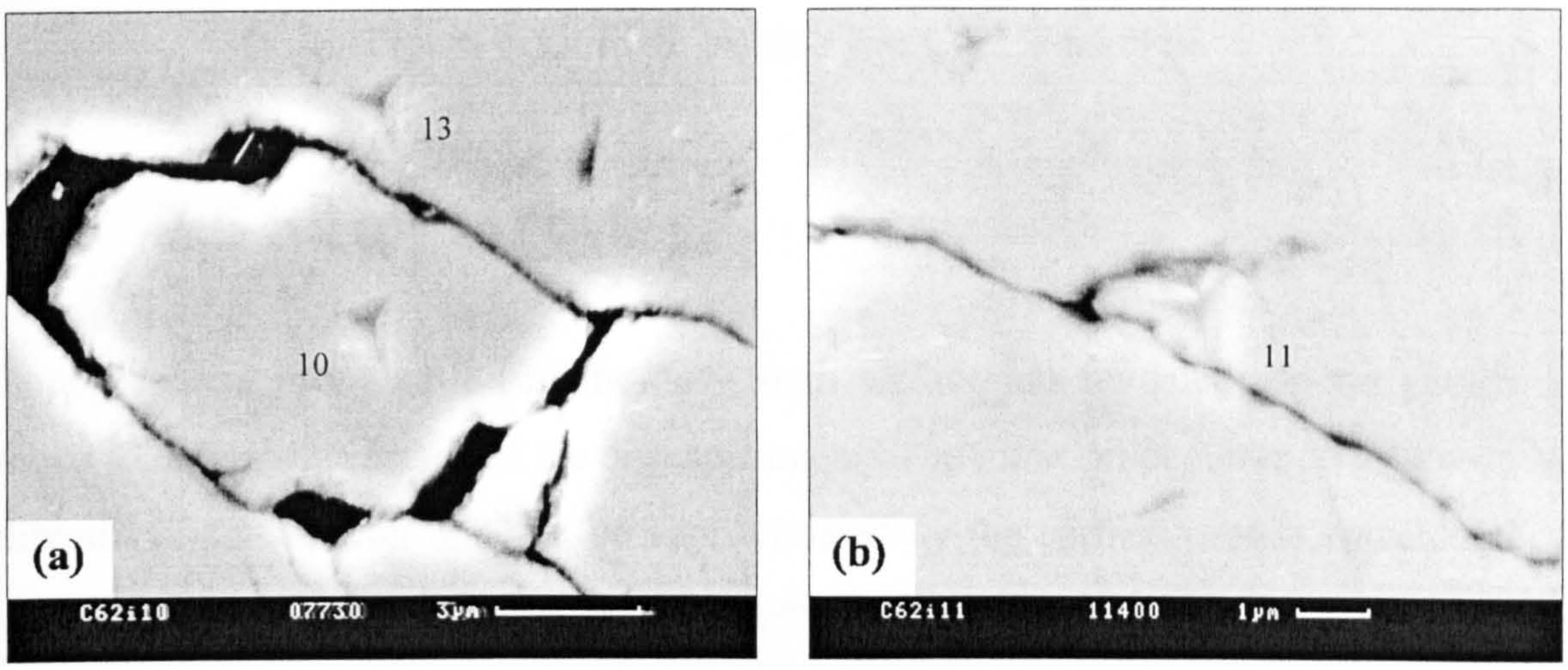


Figure 7-23. SEM images of the indentations 10 in (a) and 11 in (b).

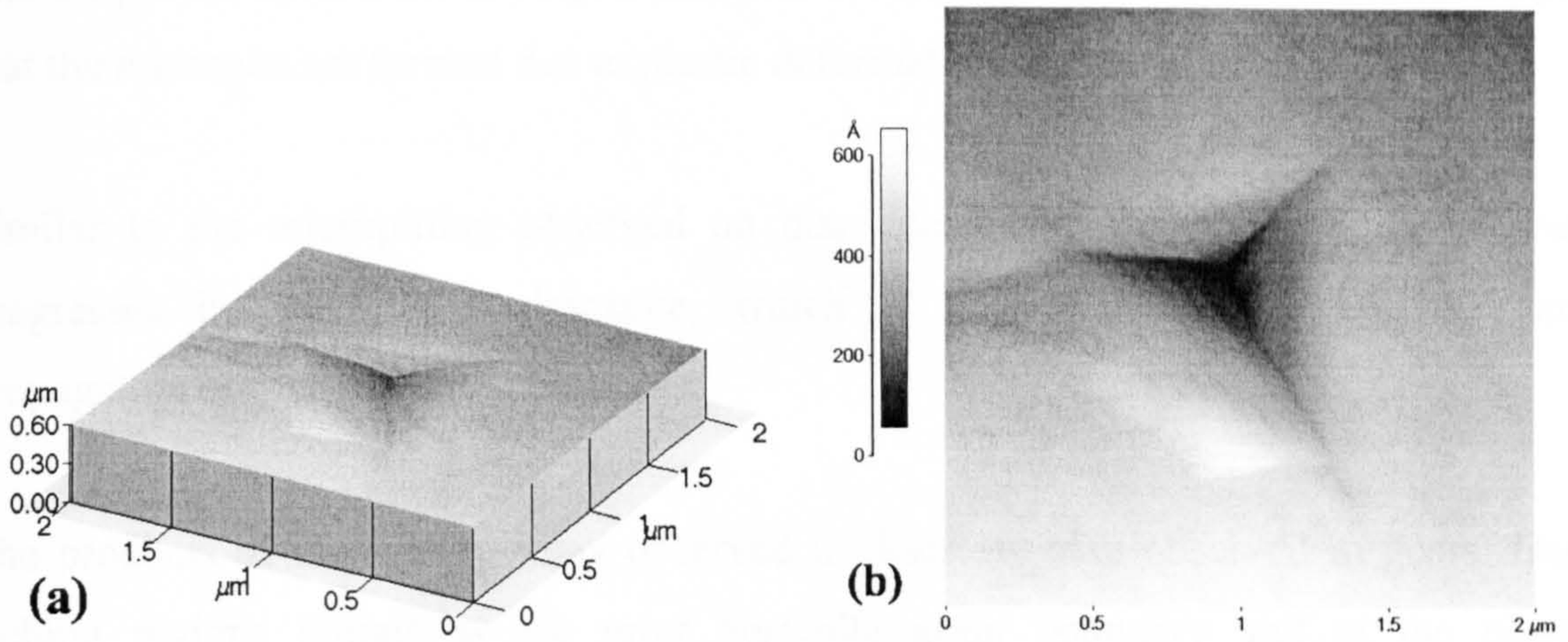


Figure 7-24. AFM images of indentation 11. (a) Three dimensional view. (b) Top view.

The micrograph from Figure 7-25 is a good example for the phase transformations that take place in gears. The figure depicts three plastic deformation regions, which appear surrounded by dark etching regions.

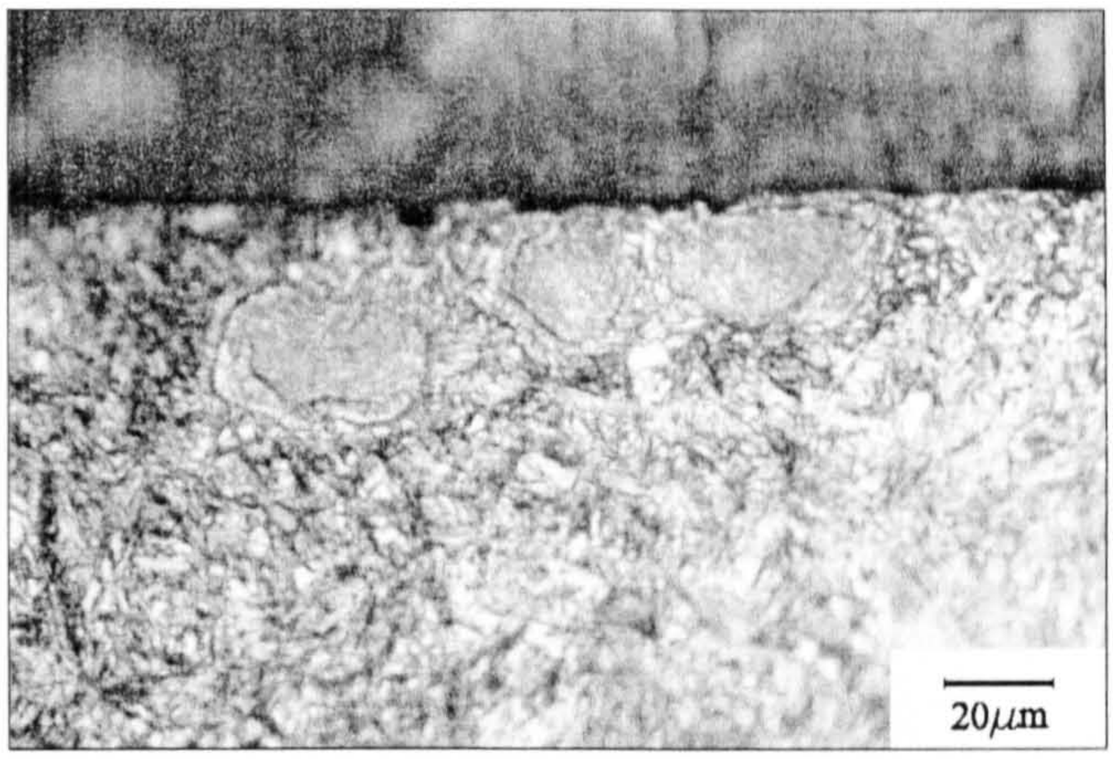


Figure 7-25. PDR grains. Gear C62. Nital etch.

7.5 Summary

Topographical investigation of the gear tooth surface has revealed aspects similar to those found on the surface of the disc specimens. The micropit occurrence follows a bi-directional distribution that can be approximated by the surface profile wavelength in two orthogonal directions: parallel and perpendicular to the surface lay.

The shape and size of the micropits and also the location of the surface cracks suggest that the micropits are formed due to plastic deformation of asperities.

Similar to the micropitting observed on disc specimens, the micropitting in gears progresses by micropit coalescence, which is explained by the initiation and propagation of multiple crack branches.

The products of martensite decay observed in discs are also observed in gears. Dark etching regions initiate at the prior austenite grain boundary and at the plastic deformation region boundary. The white etching bands do not occur as uniformly as in discs but they are randomly distributed in the near surface regions. This is explained by the variation of the direction and speed of sliding.

The mechanism of micropitting in discs and gears is the same and it is based on the phase transformations. Due to severe plastic deformation of asperities, microcracks

initiate at the surface, at the point of intersection between the surface and the boundary of the plastic deformation region. It propagates along the boundary until it reaches the surface again. When the crack reaches the surface a micropit is being formed.

The boundary of the plastic deformation region consists of dark etching grains, which are softer than the plastic deformation region and the zones located below it.

~ Chapter 8 ~

FACTORS INFLUENCING MICROPITTING

8 FACTORS INFLUENCING MICROPITTING

8.1 General

The results presented and discussed in the previous chapters are used to assess the factors that influence the initiation and propagation of micropitting. Since there are two sets of experimental conditions for each test, i.e. the *initial* and *final conditions*, which were given in Table 5-3, the fractional factorial design has been applied twice. At first the initial conditions were used in order to assess the influence of the factors on the initiation of micropitting and then the final conditions in order to analyse the influence of factors on the propagation of micropitting.

8.2 The factors of influence

The factors used in the experimental design, as well as their levels were presented in section 4.3 (see Table 4.2). The selection of the factors considered has been treated in detail in section 3.7. In order to assess their effects and to determine the regression equations these factors must be represented numerically. In the following sections are given the considerations used to assign the numerical values to each factor.

8.2.1 Steel grade

The fatigue resistance of the two steels used in micropitting experiments depends on the chemical composition (carbon content, non-metallic inclusions, alloying elements) as well as the microstructure (martensite, retained austenite). The properties that are most relevant to fatigue behaviour are hardness and elastic modulus. However, these properties should be determined at the surface, which is a difficult task because of the

surface roughness (see section 5.9). It would be improper to use a hardness value determined by the conventional Vickers method because the steel microstructure is not homogeneous. It has been reckoned that the most suitable property, which characterises the steel in the contact region, is the *elastic contact modulus* (or *reduced modulus*), E' given by the equation 2-16.

8.2.2 Surface lay

It has been shown in the previous chapters that the formation of micropits is related to the asperity contact. The micropitting mechanism proposed in this work is based on the plastic deformation of asperities. It follows that the surface roughness is one of the most important factors that influence micropitting. The investigations carried out on the affected surfaces, of both discs and gear teeth have revealed a periodicity of the micropits, which can be described by the surface roughness wavelength in two orthogonal directions. The parameter used to evaluate the influence of surface roughness is the *wavelength ratio*, Λ, defined in section 5.5 (see Table 5-5).

8.2.3 Load

The plastic deformation of asperities precedes the initiation of cracks, which will lead to the formation of micropits. The contact pressure is the factor that governs the plastic deformation of asperities. The parameter used to characterise the effect of load is the maximum contact pressure, p_0.

8.2.4 Lubricant

The most important property of a lubricant is represented by its viscosity. The viscosity of the lubricants used in this work varies with temperature and pressure as shown in Figures 4-9 & 10. However, these variations depend on the viscosity at ambient temperature and pressure, η_0. The ambient viscosity, η_0, has been used to assess the effect of lubricant on the micropitting phenomenon.

8.2.5 Temperature

The presence of the martensite decay products in both, discs and gears, which involves carbon diffusion indicates that the thermal effects cannot be neglected. The temperature within the contact is higher than the temperature of lubricant measured in the disc housing. Moreover the flash temperature arising from asperity contacts is difficult to estimate it. In the absence of a solid mathematical apparatus that could be applied to determine the flash temperature, the temperature used in the factorial design is the temperature of the lubricant inside the disc housing.

8.2.6 Speed

The entraining velocity, u, is commonly used in the calculations which refer to elastohydrodynamic contacts. Because all the measurements have been carried out on the driven disc, the entraining velocity considered in the statistical calculations is that of the driven disc.

8.2.7 Slide-to-roll ratio

Sliding has a major effect on the friction coefficient. The friction coefficient specific to each test has been given in Table 5-5. The slide-to-roll ratio has been calculated using formula 2-6.

8.3 Micropitting initiation

The initial contact conditions have been used to study the influence of the given factors on the number of cycles N_0, necessary for micropitting to initiate. N_0 has been determined for each experiment in section 5-6. It is considered in this work that micropitting had been initiated when the micropits occupy 1.5 % of the disc surface. The initial conditions have been also used to analyse their effect on the specific film thickness (lambda ratio), λ. The matrix of the factorial design is given in Table 8-1.

Table 8-1. The factorial design matrix for micropitting initiation.

Run	Factors							Responses	
	E' (GPa)	Λ	p_0 (GPa)	η_0 (Pas)	T(°C)	u(m/s)	S	N_0	λ
1	106.66	0.335	1.8	0.169	60	3.76	0.33	$1.48 \cdot 10^5$	0.38
2	111.28	2.547	2.2	0.169	60	3.76	0.23	$4.55 \cdot 10^4$	0.66
3	106.66	2.547	1.8	0.268	60	3.13	0.23	$4.72 \cdot 10^5$	0.34
4	111.28	2.547	1.8	0.169	100	3.13	0.33	$1.05 \cdot 10^3$	0.34
5	111.28	0.335	1.8	0.268	100	3.76	0.23	$1.12 \cdot 10^5$	0.27
6	106.66	2.547	2.2	0.268	100	3.76	0.33	$1.17 \cdot 10^4$	0.14
7	106.66	0.335	2.2	0.169	100	3.13	0.23	$4.69 \cdot 10^4$	0.38
8	111.28	0.335	2.2	0.268	60	3.13	0.33	$6.25 \cdot 10^3$	0.48

8.3.1 The critical number of cycles, N_0

Figure 8-1 shows the main effects plots when the response is N_0. The central line represents the mean of all the values of the response (N_0) in the experiment. The points situated above the mean line show the mean of all runs where the factors are at the high setting and the points below the central line show the mean of all runs where the factors are at the low setting. The main effect of a selected factor is the difference between the low setting and the high setting on the graph.

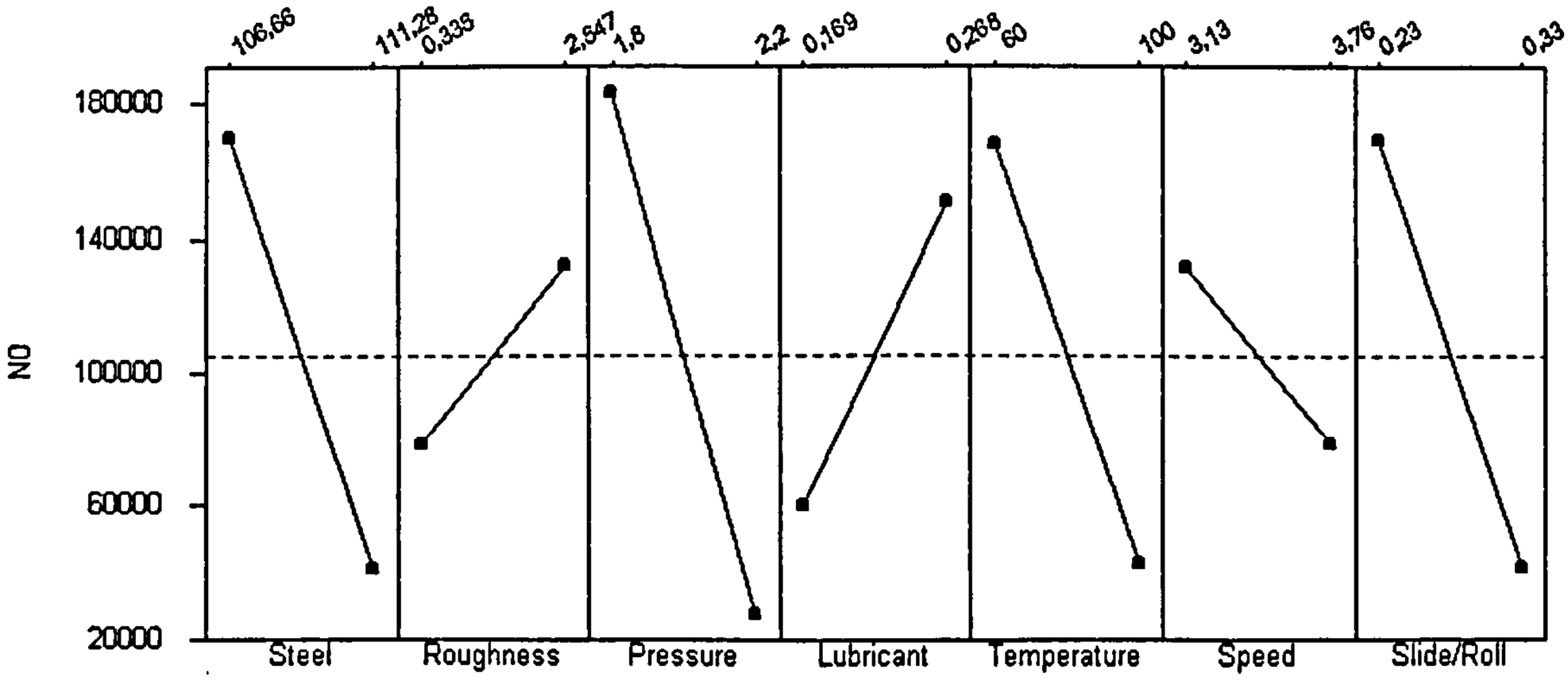


Figure 8-1. The main effects on micropitting initiation. N_0 is the number of cycles after micropitting occupies 1.5% of the surface.

To calculate main effects, the Minitab software subtracts the mean response at the low level of the factor from the mean response at the high level of the factor.

The biggest effect on the number of cycles, N_0, necessary for micropitting to initiate is that of the contact pressure. The slope of the effect line shows that micropitting initiates considerably faster when the contact pressure is high. Although the pressure appears to affect N_0 more than the other factors it is important to analyse the interactions. The interaction plots when the response is N_0 are shown in Figure 8-2. An interaction plot shows the impact that changing the settings of one factor has on another factor. Because an interaction can magnify or diminish main effects, evaluating interactions is extremely important.

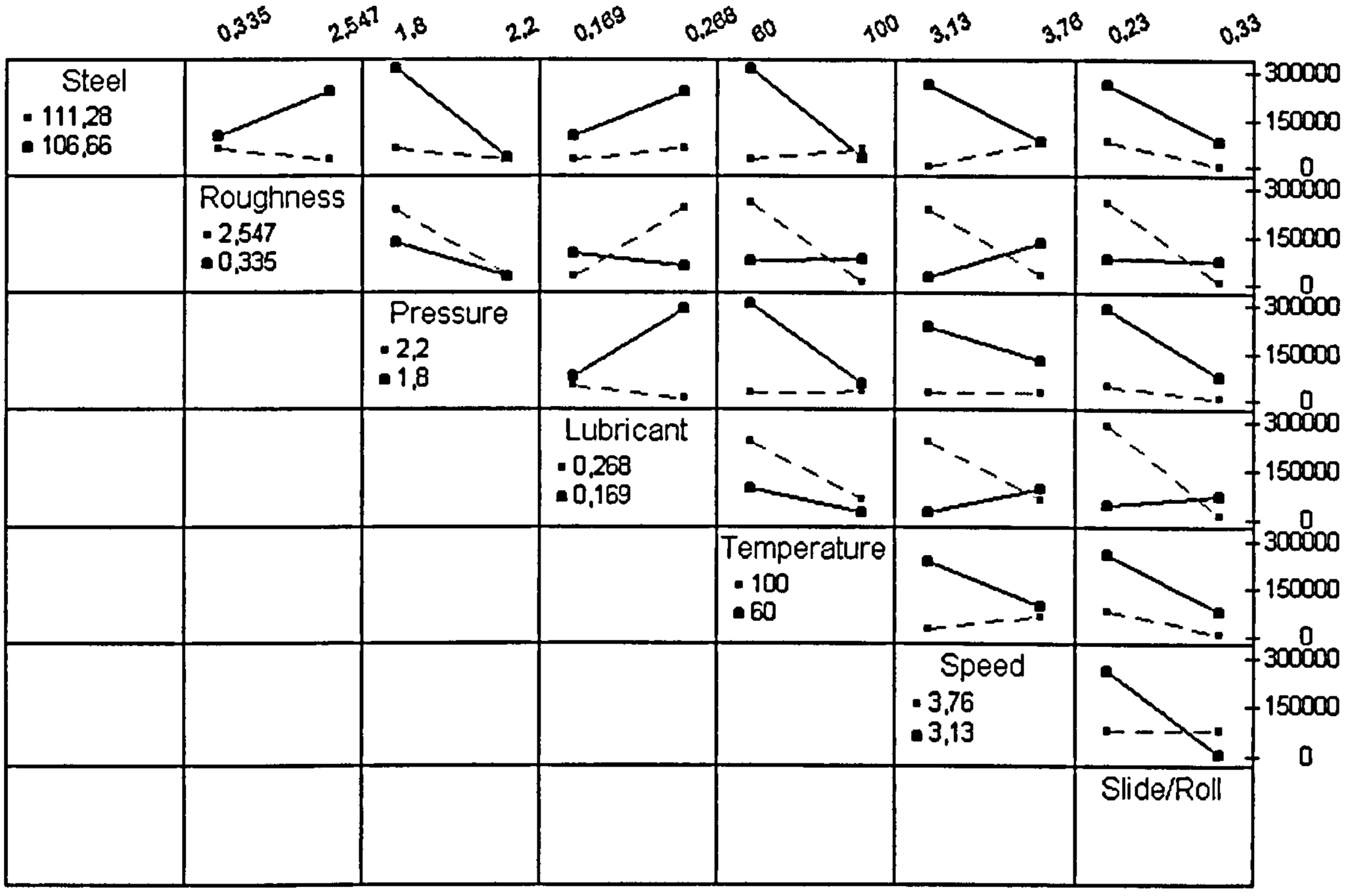


Figure 8-2. The interactions plots for micropitting initiation.

The interaction plots show that the number of cycles, N_0, necessary for micropitting to initiate for high contact pressure (p_0 = 2.2 GPa) is almost independent by the other factors. The only significant interaction is that with surface roughness. This observation is very important. At high contact pressures the means to avoid micropitting are limited and, under the experimental conditions used in this research the only way to retard micropitting initiation is represented by a surface finish improvement. At the low setting

for contact pressure (p_0 = 1.8 GPa) micropitting initiation is strongly affected by the other factors. N_0 increases when EN36 steel and mineral oil are used, when the wavelength ratio, Λ, the temperature, T, the speed, u, and the slide-to-roll ratio, S, decrease.

The second biggest influence on the initiation of micropitting is the elastic contact modulus of the steel. Micropitting initiates earlier on the specimens manufactured from SAE 8620 steel, which is harder than EN36 (higher carbon content) and has a bigger elastic contact modulus (E' = 111.28 MPa). This contradictory result is probably caused by other material properties, which have not been considered. The interaction plots show that for SAE 8620 steel, N_0 varies insignificantly when the other factors vary. In conclusion, the SAE 8620 steel is more prone to micropitting. The initiation of micropitting in the EN36 steel is influenced by all the other factors. Better resistance to micropitting initiation is provided by high wavelength ratio, low contact pressure, the synthetic lubricant, low temperature, low speed and low slide-to-roll ratio.

Slide-to-roll ratio also has a significant influence on micropitting initiation. Lower the ratio higher the number of cycles for micropitting to occur. Slide-to-roll ratio has no effect when the wavelength ratio is low and also at high speed.

Table 8-2. The size of effects. Response is N_0.

Factor	Size of effect
Steel	-128450
Roughness	54275
Pressure	-155675
Lubricant	90125
Temperature	-125025
Speed	-52250
Slide/Roll	-127350

The effect of temperature is as expected, at low temperatures the occurrence of micropitting is retarded. Temperature has no influence at high contact pressure.

The best micropitting performance is provided by mineral lubricant. Both lubricants act better at high temperatures in the sense that the initiation of micropitting is delayed. At

high temperatures the additives are activated and they form a protective layer on the disc surface.

The influence of surface roughness and speed are the smaller and they can be neglected.

8.3.2 The specific film thickness, λ

It has been shown in the previous chapters that micropitting is caused by plastic deformation of asperities. The parameter that estimates the degree of metallic contact (asperity contact) is the specific film thickness or lambda ratio, λ. The main effects plots and the interaction plots when the response is the specific film thickness, λ, are shown in Figure 8-3 and 8-4, respectively. The effects of surface roughness and speed are not significant. The biggest influence on lambda ratio is the effect of temperature. At high temperature the lubricant film thickness decreases and the collision between asperities becomes more probable. When the temperature is high, lambda decreases considerably with both types of surface roughness and lubricants. The decrease is accentuated by high speeds. Slide-to-roll ratio has little effect on the specific film thickness at elevated temperature and λ remains low. An important increase of λ can be achieved at lower slide-to-roll ratios.

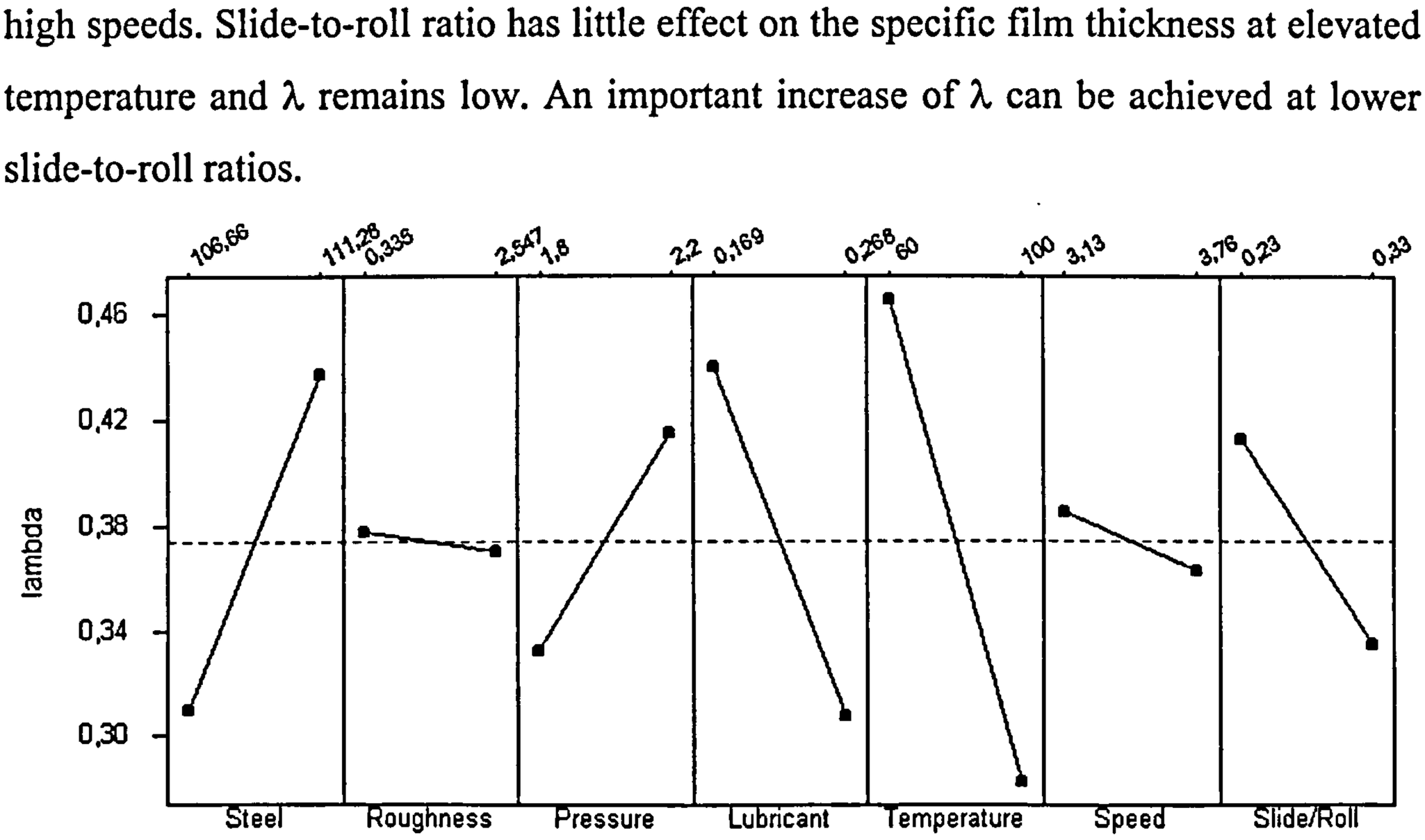


Figure 8-3. The main effects on the lambda ratio during the period of micropitting initiation.

The second biggest effect on λ is exerted by lubricant viscosity. Better lubrication conditions (higher λ) are achieved when mineral oil is used. Lubricant has no effect on

λ if the surface lay is transverse but the synthetic lubricant produces very low specific film thickness. λ decreases at elevated temperature and for both temperature levels the use of mineral oil results in higher λ values. A significant interaction takes place between lubricant and speed. Lambda ratio increases at high speed when the lubricant is mineral oil and reduces when the lubricant is OEP-80.

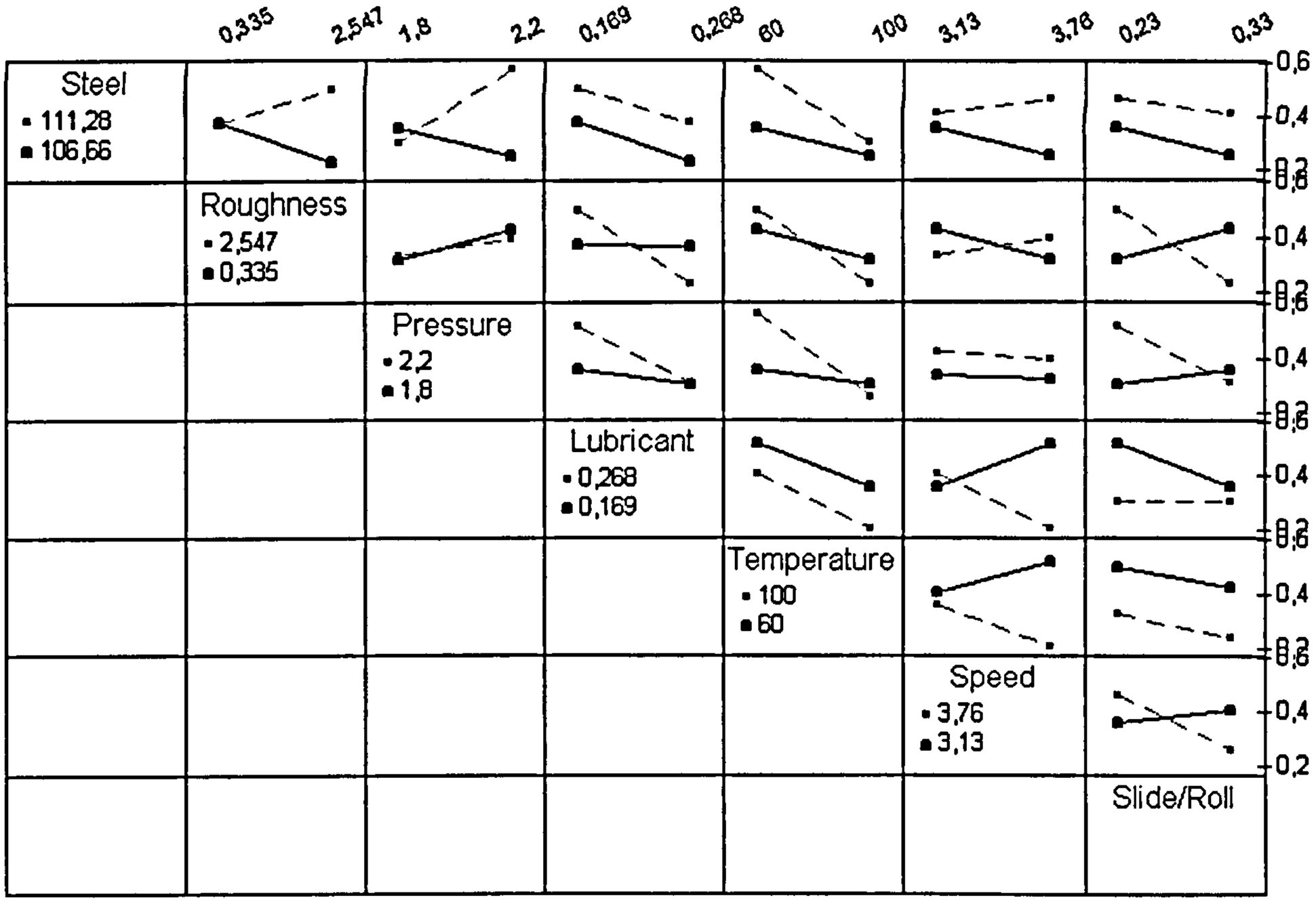


Figure 8-4. The interactions plots for the lambda ratio, during the initiation period.

Table 8-3. The size of effects. Response is λ.

Factor	Size of effect
Steel	0.1275
Roughness	-0.0075
Pressure	0.0825
Lubricant	-0.1325
Temperature	-0.1825
Speed	-0.0225
Slide/Roll	-0.0775

The overall conclusion is that mineral oil produces higher specific film thickness.

The influence of steel type is also significant; lambda ratio is lower for EN36 steel. The other effects are small and they can be neglected.

8.4 Micropitting propagation

The variation of the surface roughness parameters leads to changes in the operating conditions. The factors affected by these changes are the wavelength ratio, Λ, and the maximum contact pressure, p_0.

Table 8-4. The factors and their levels used to study the micropitting propagation.

Run	Factors						
	E' (GPa)	Λ	p_0 (GPa)	η_0 (Pas)	T (°C)	u (m/s)	S
1	106.66	0.642	1.5	0.169	60	3.76	0.33
2	111.28	1.67	1.8	0.169	60	3.76	0.23
3	106.66	1.67	1.5	0.268	60	3.13	0.23
4	111.28	1.67	1.5	0.169	100	3.13	0.33
5	111.28	0.642	1.5	0.268	100	3.76	0.23
6	106.66	1.67	1.8	0.268	100	3.76	0.33
7	106.66	0.642	1.8	0.169	100	3.13	0.23
8	111.28	0.642	1.8	0.268	60	3.13	0.33

Table 8-5. The responses used to study the micropitting propagation.

Run	Responses					
	dM/dN	λ	$p \cdot 10^{-5}$ (μm/cycle)	$d \cdot 10^{-5}$ (μm/cycle)	$w \cdot 10^{-5}$ (μm/cycle)	α (°)
1	$6.80 \cdot 10^{-5}$	0.35	1.25	51	140	30
2	$3.87 \cdot 10^{-5}$	0.35	0	19	0	31
3	$8.77 \cdot 10^{-6}$	0.57	0.34	4	0	60
4	$3.06 \cdot 10^{-5}$	0.24	1.11	5	0	28
5	$3.21 \cdot 10^{-5}$	0.33	0	1	0	20
6	$8.55 \cdot 10^{-5}$	0.17	2.22	102	352	42
7	$2.83 \cdot 10^{-5}$	0.31	1.38	38	202	24
8	$1.74 \cdot 10^{-5}$	0.65	0	3	0	16

The *final conditions* are used to study the stage of micropitting propagation. The responses used to asses the influence of the factors on the propagation of micropitting are: *the micropitting propagation rate*, d*M*/d*N* (the slopes of the micropitting propagation curves), *the specific film thickness*, λ, *the phase transformation factors* (i.e., *p* the plastic deformation regions factor, *d*, the dark etching regions factor and *w* the white etching bands factor) and *the angle of crack propagation*, α. The factors and their levels corresponding to each experiment are given in Table 8-4 and the response columns in Table 8-5.

8.4.1 Micropitting propagation rate, d*M*/d*N*

The main effects plots when the response is the micropitting propagation rate, d*M*/d*N*, are shown in Figure 8-5. The effect of speed, slide-to-roll ratio and material are much higher than the other effects and they are significant. Micropitting propagates with high rates at high speed and slide-to-roll ratio.

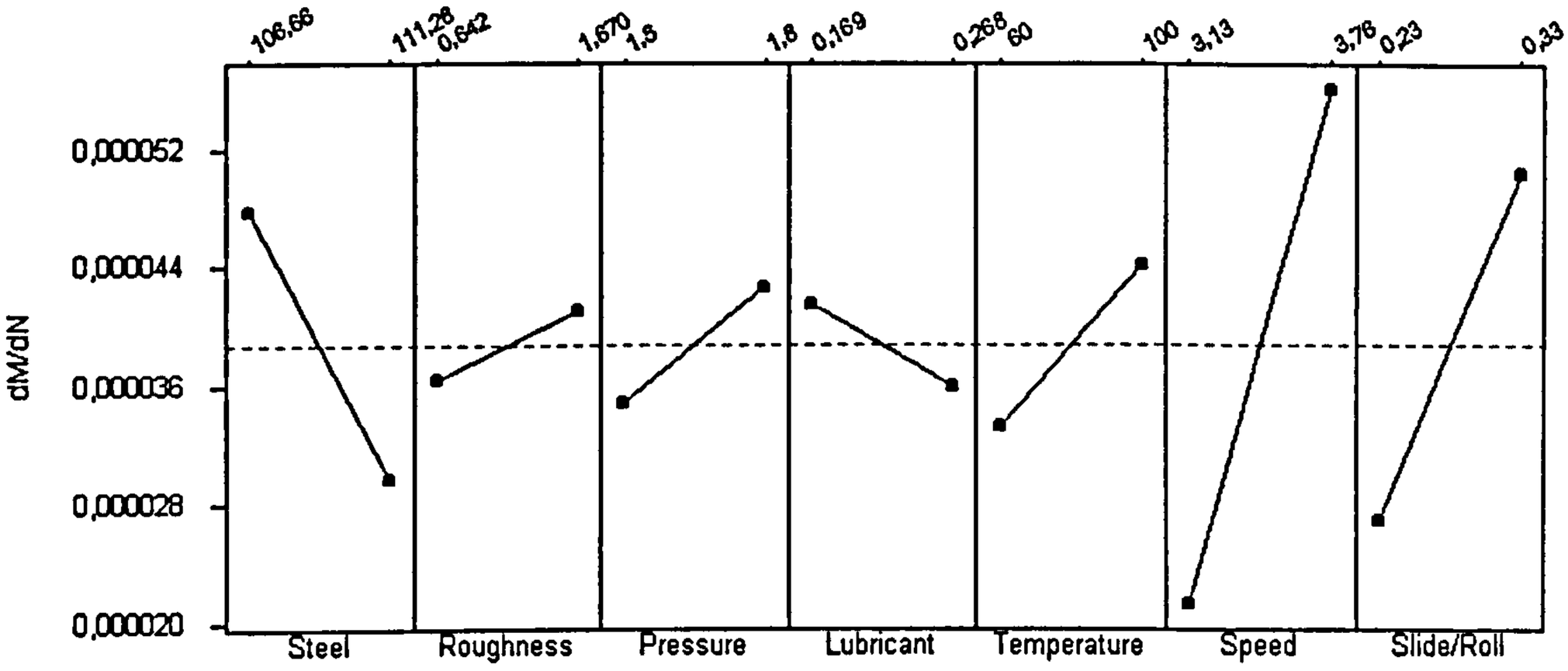


Figure 8-5. The main effects on micropitting initiation. d*M*/d*N* represents the micropitting propagation rate.

Although, the micropitting initiation is delayed on EN36 steel, as seen in Figure 8-1, the propagation rate is much bigger than for SAE 8620 steel.

From the interaction plots (Figure 8-6) all the interactions which involve speed are significant. At high speed micropitting propagates with high rate regardless the values

of the other factors. This implies that little can be done to stop the progression of micropitting when the operating speed is high.

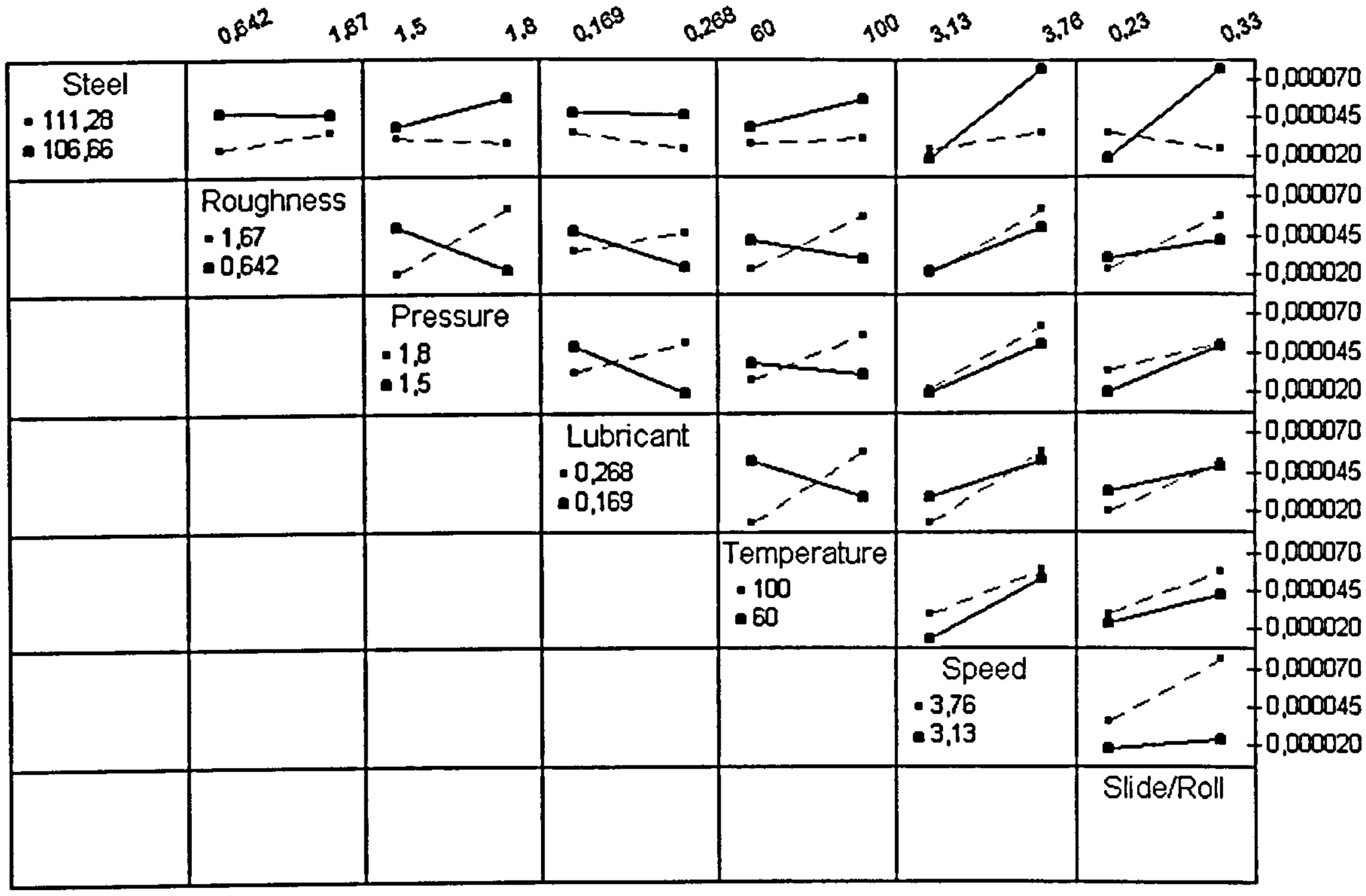


Figure 8-6. The interactions plots for micropitting propagation.

The propagation rate increases dramatically when high speed is coupled with high slide-to-roll ratio. The effect of increasing speed and slide-to-roll ratio is more pronounced on EN36 steel.

Table 8-6. The size of effects. Response is d*M*/d*N*.

Factor	Size of effect
Steel	-0.000018
Roughness	0.000004
Pressure	0.000008
Lubricant	-0.000005
Temperature	0.000011
Speed	0.000035
Slide/Roll	0.000023

8.4.2 Specific film thickness, λ

It has been shown that, during the initiation period, the biggest effects on specific film thickness are those of temperature and lubricant. The main effects plots for lambda ratio during the propagation period (Figure 8-7) show that the biggest effect remains the effect of temperature. Lubricant still has a significant influence but the roles are reversed; smaller values for λ are given by mineral oil. However, during the propagation period the second biggest effect is that of speed.

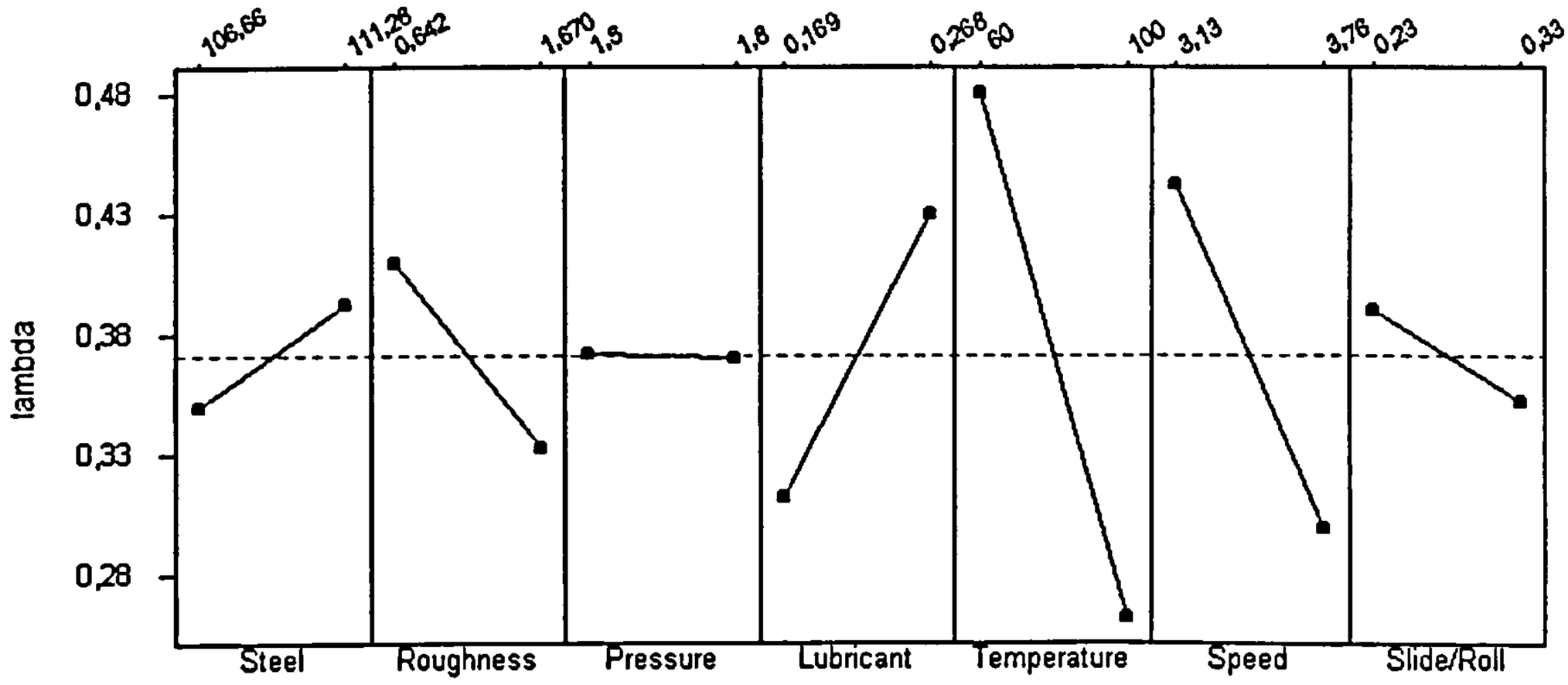


Figure 8-7. The main effects on the lambda ratio during the period of micropitting propagation.

The interaction plots for λ during the propagation period are shown in Figure 8-8. It can be observed that better lubrication conditions (higher λ) are achieved at high speed and slide-to-roll ratio at both temperatures. Considering that micropitting propagates with maximum rates at high speed and slide-to-roll ratio this result may appear unexpected. Although the number of asperity contacts is reduced (higher λ) the progression of micropitting is accelerated. This suggests that, once initiated, micropitting propagation cannot be reduced by improving lubrication conditions. It is interesting to note that the two lubricants act differently as a function of temperature. The specific film thickness decreases with temperature for mineral oil and increases for OEP-80.

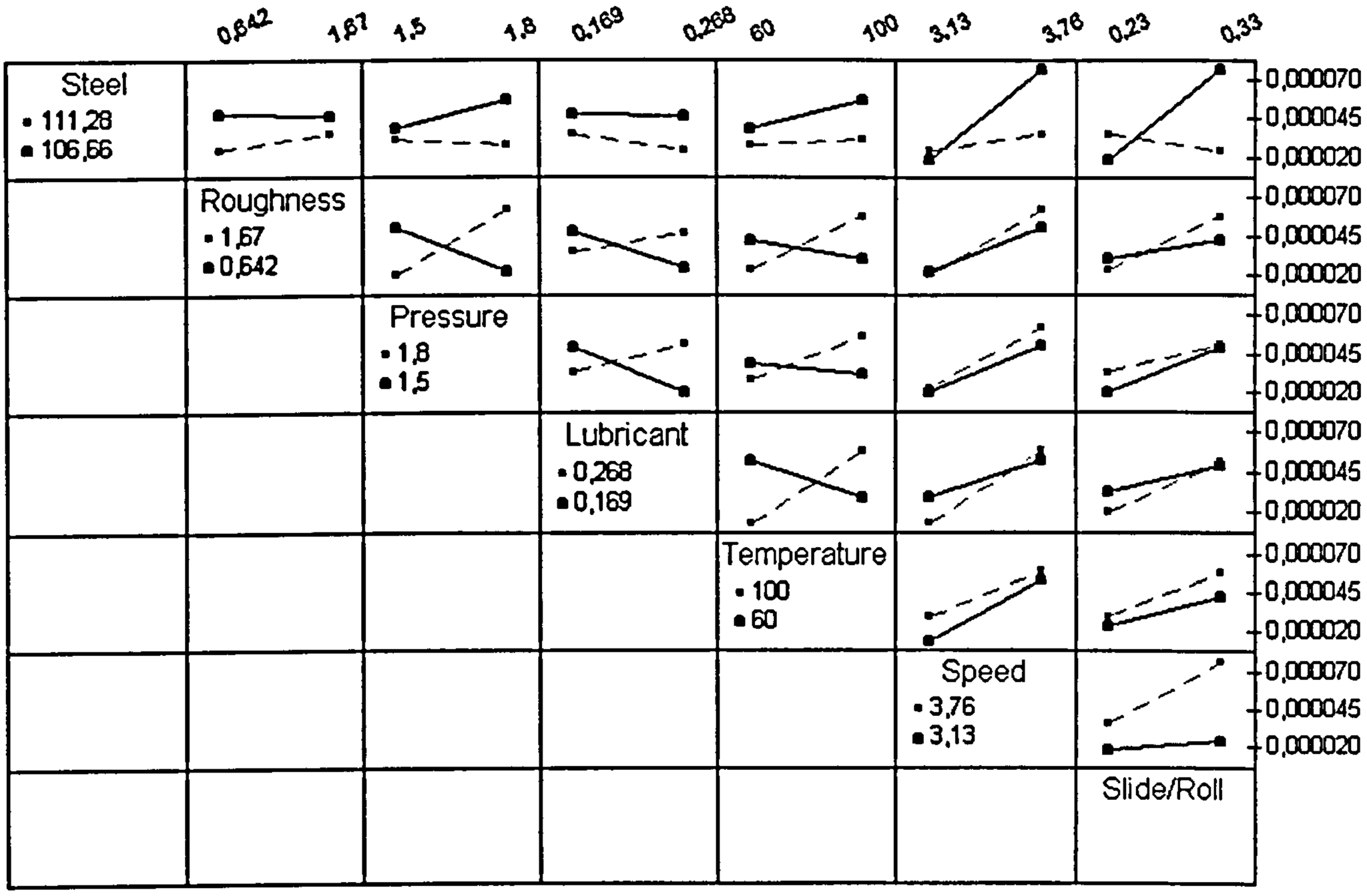


Figure 8-8. The interactions plots for the lambda ratio, during the propagation period.

Table 8-7. The size of effects. Response is λ.

Factor	Size of effect
Steel	0.0425
Roughness	-0.0775
Pressure	-0.0025
Lubricant	0.1175
Temperature	-0.2175
Speed	-0.1425
Slide/Roll	-0.0375

8.4.3 Crack propagation angle, α

Figure 8-9 shows the main effects plots when the response is the angle of crack propagation, α. The angle of propagation depends mostly on steel and surface roughness wavelength ratio, Λ. The effects produced by the other factors are attenuated

by these effects. The angle of propagation is smaller for low wavelength ratios (longitudinal surface lay).

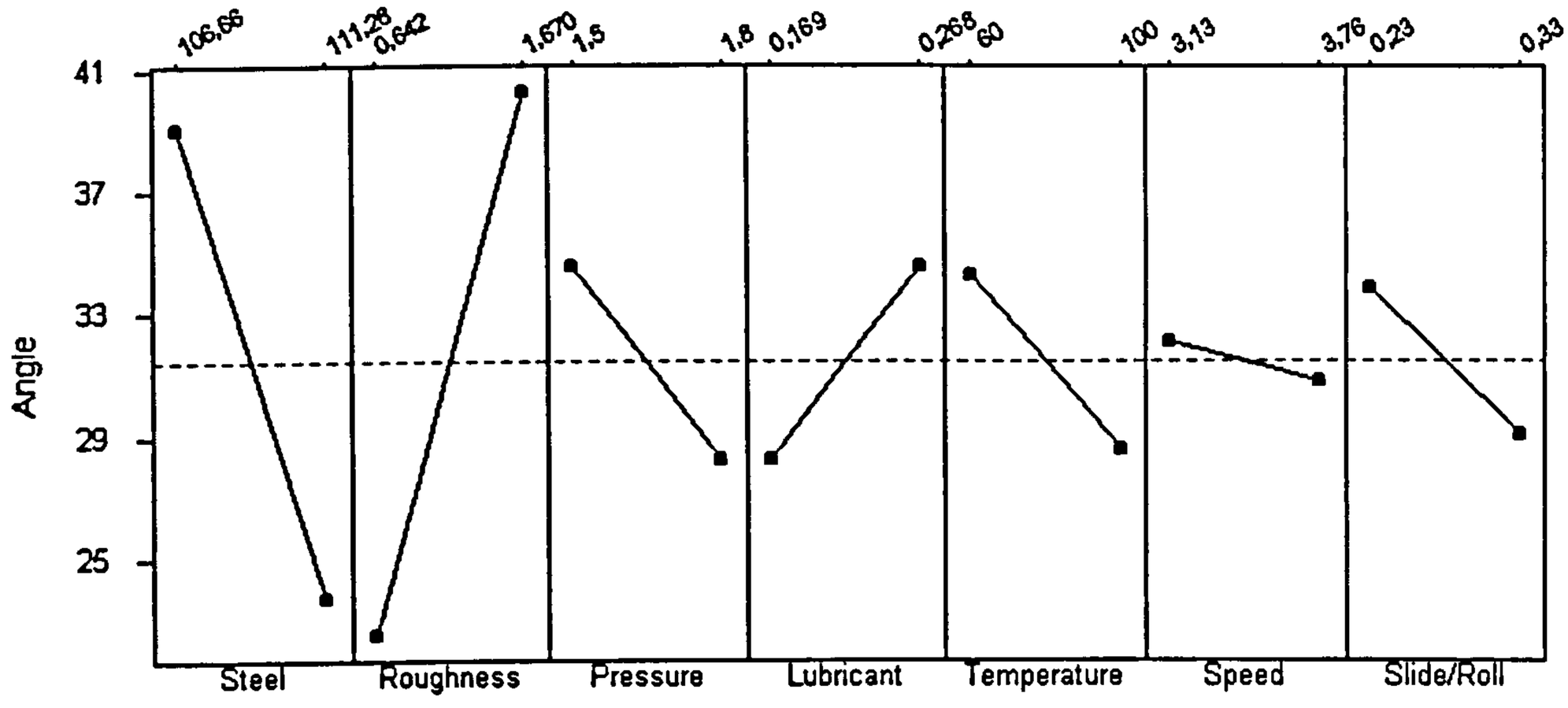


Figure 8-9. The main effects on the crack angle.

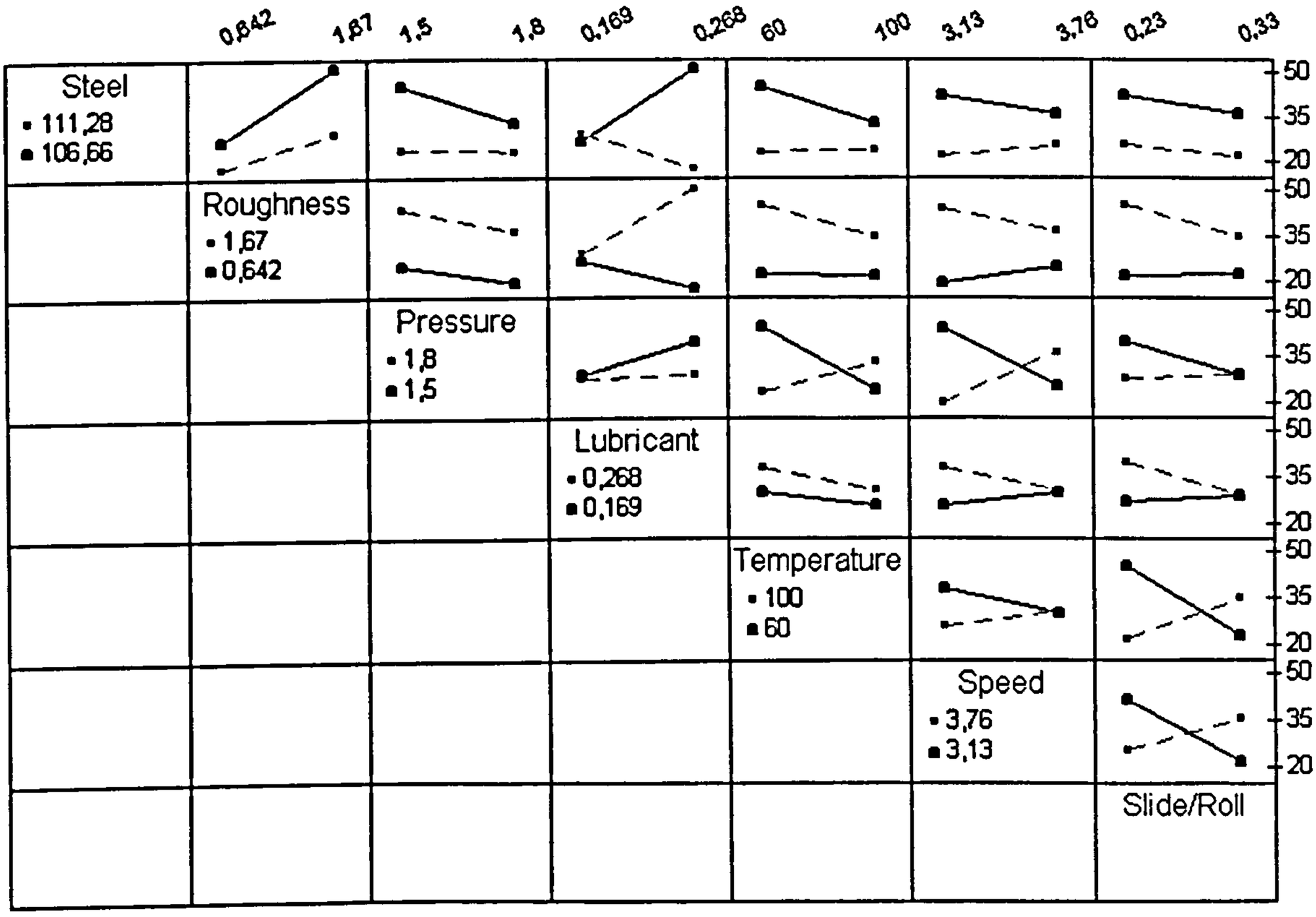


Figure 8-10. The interactions plots for the crack angle.

The interaction plots for α are shown in Figure 8-10. The interaction between wavelength ratio and lubricant is significant. The angle of inclination is not affected by

mineral oil but increases significantly when OEP-80 is used to lubricate surfaces with high wavelength ratio. Cracks propagate at smaller angles in SAE 8620 steel than in *EN36 steel.*

Table 8-8. The size of effects. Response is α.

Factor	Size of effect
Steel	-15.250
Roughness	17.750
Pressure	-6.250
Lubricant	6.250
Temperature	-5.750
Speed	-1.250
Slide/Roll	-4.750

8.4.4 Phase transformations

The phase transformations associated with micropitting have been treated in Chapters 6 & 7. It has been also shown that the two phenomena are related to each other. In order to study the effects that the seven factors have on phase transformations, the factorial calculations have been done using the phase transformation factors, *p* - the PDR factor, *d* - the DER factor, and *w* - the WEB factor (they were defined in Chapter 6) as responses.

8.4.4.1 Plastic deformation layer

The plastic deformation zones form more extensively in samples manufactured from EN36 steel. The biggest effect in the main effects plots (Figure 8-11) is the effect of elastic contact modulus of the steel. The second biggest effects are exerted by temperature and slide-to-roll ratio. The fact that thermal effect has a high level of significance can be explained by the flash temperatures.

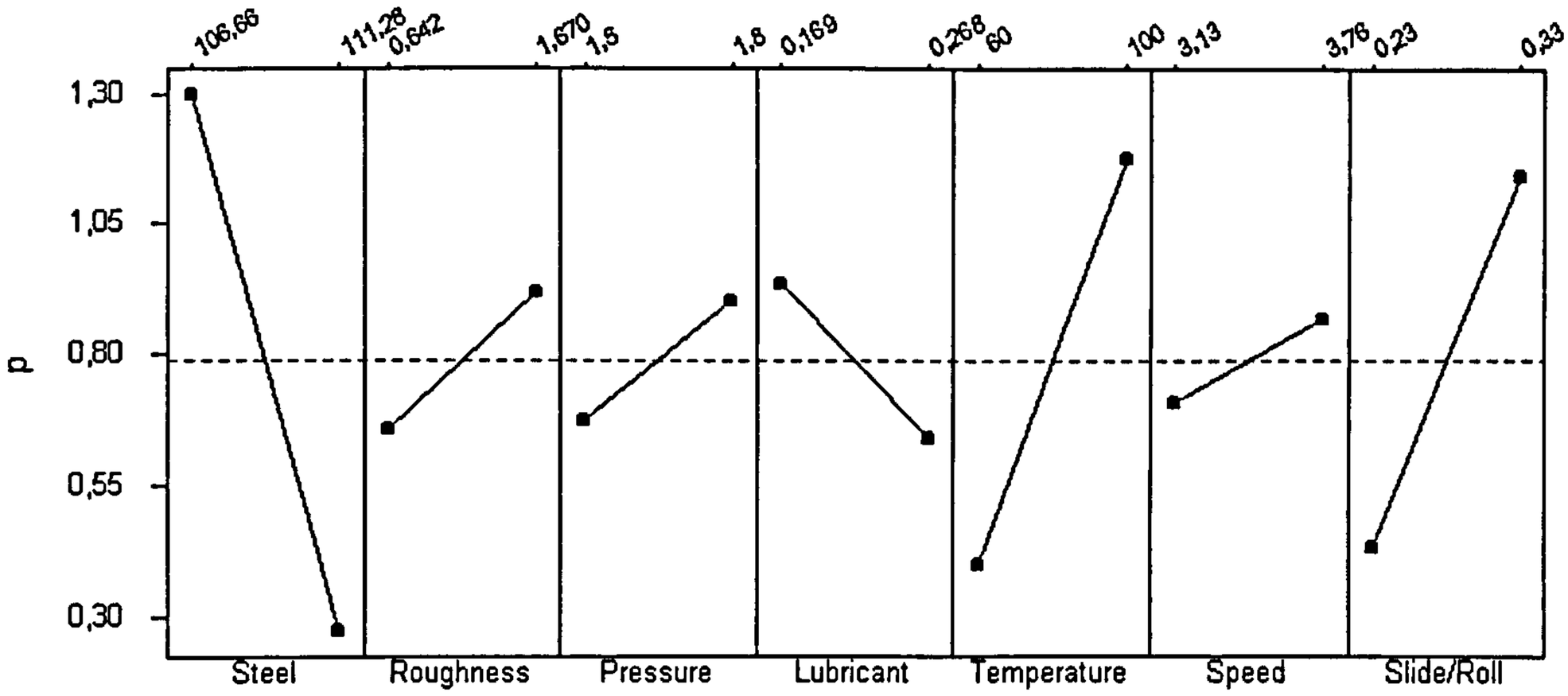


Figure 8-11. The main effects on the development of the plastic deformation regions.

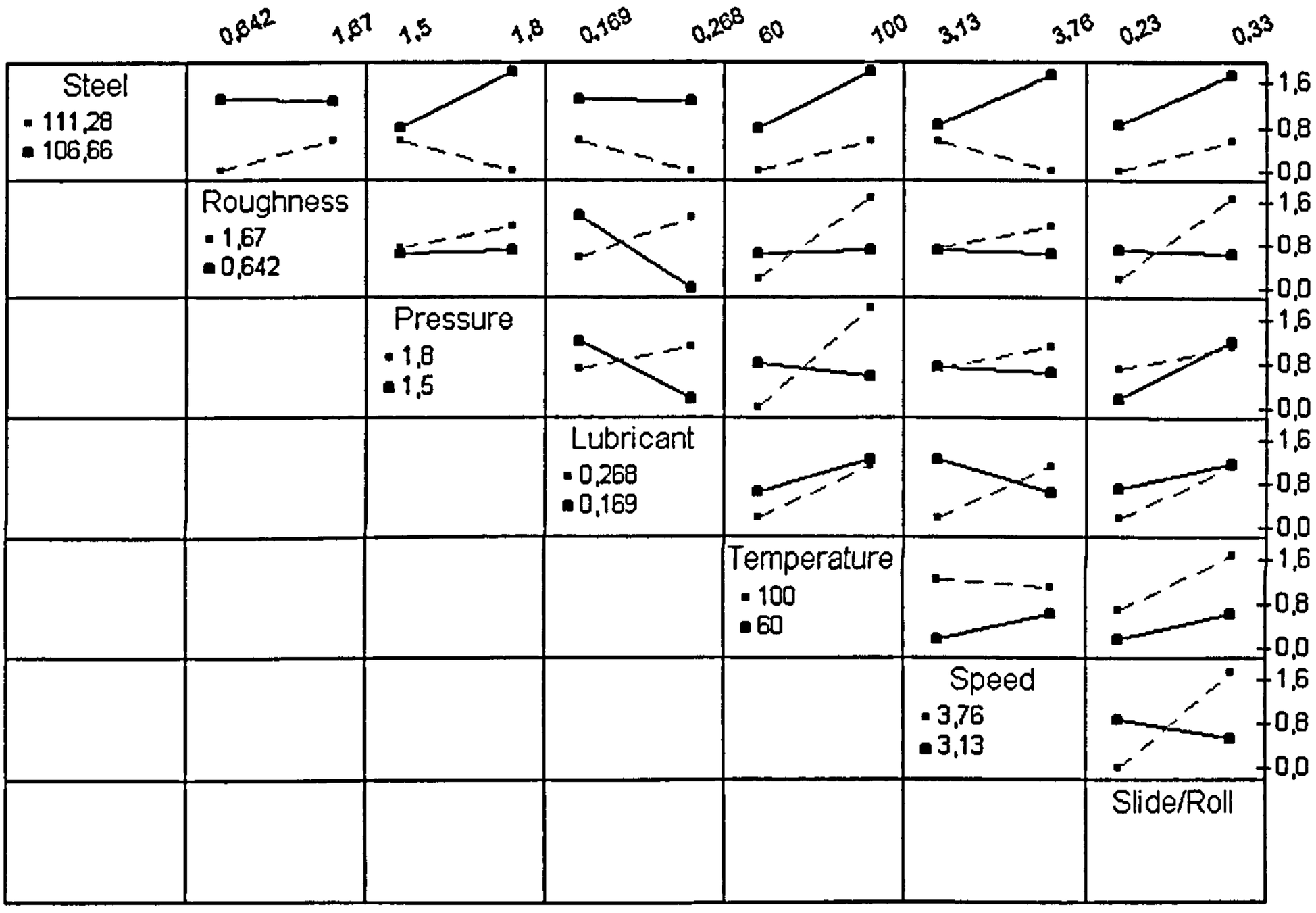


Figure 8-12. The interactions plots for plastic deformation regions.

Table 8-9. The size of effects. Response is *p*.

Factor	Size of effect
Steel	-1.0200
Roughness	0.2600
Pressure	0.2250
Lubricant	-0.2950
Temperature	0.7800
Speed	0.1600
Slide/Roll	0.7150

8.4.4.2 Dark etching regions

From the main effects plots shown in Figure 8-13 results that the development of dark etching regions depends strongly on the steel. The effect of steel is the biggest and the dark etching regions develop preferentially in EN36 steel.

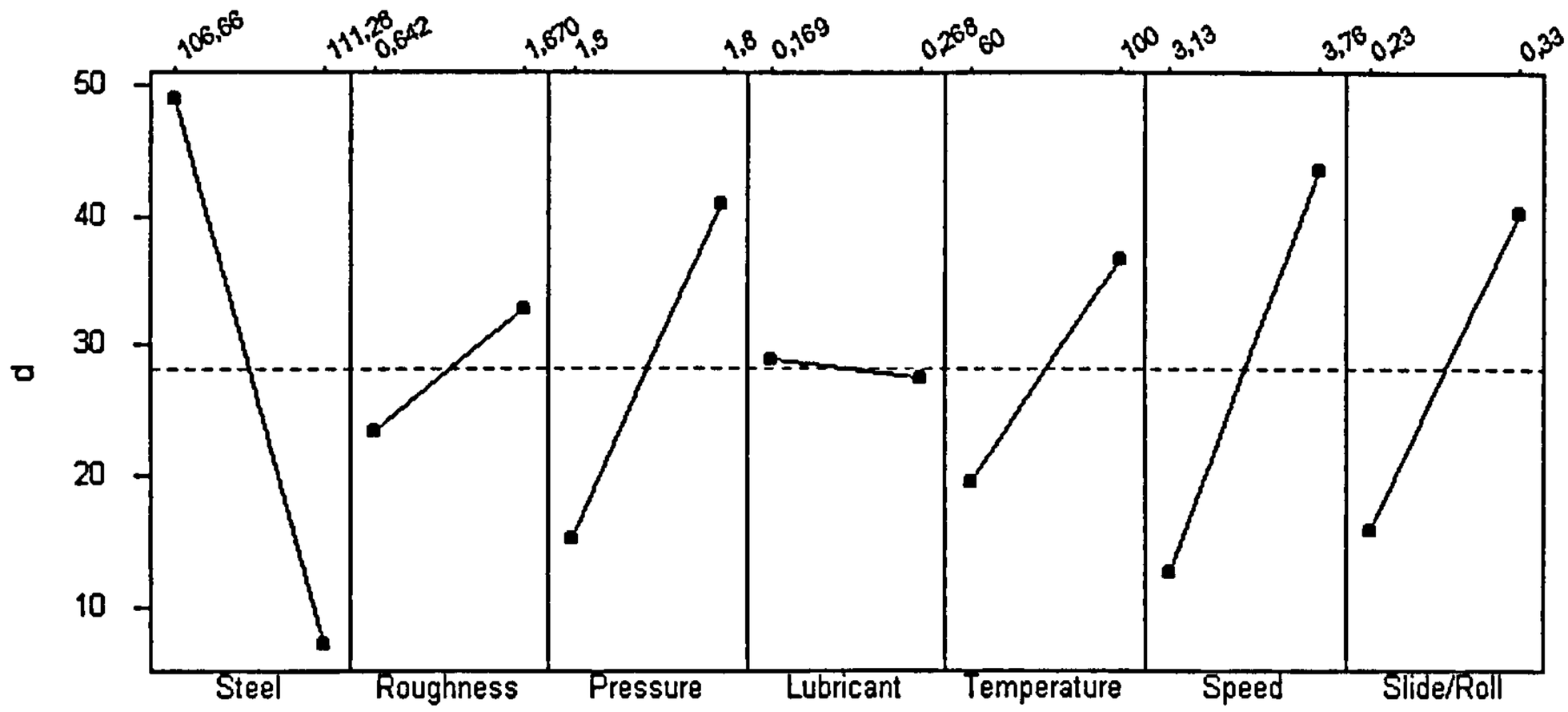


Figure 8-13. The main effects on the development of the dark etching regions.

The next significant effects are exerted by speed, contact pressure, slide-to-roll ratio and, temperature, in this order. The factors that have the biggest influence on micropitting propagation, namely speed and slide-to-roll ratio have also a big influence on dark etching regions development. This observation confirms that micropitting and martensite decay are related to each other. The DER factor does not vary significantly for SAE 8620 steel, regardless the value of the other influencing factors but the effect of

speed, slide-to-roll ratio, temperature and pressure is pronounced for EN36 steel (see the interaction plots in Figure 8-14).

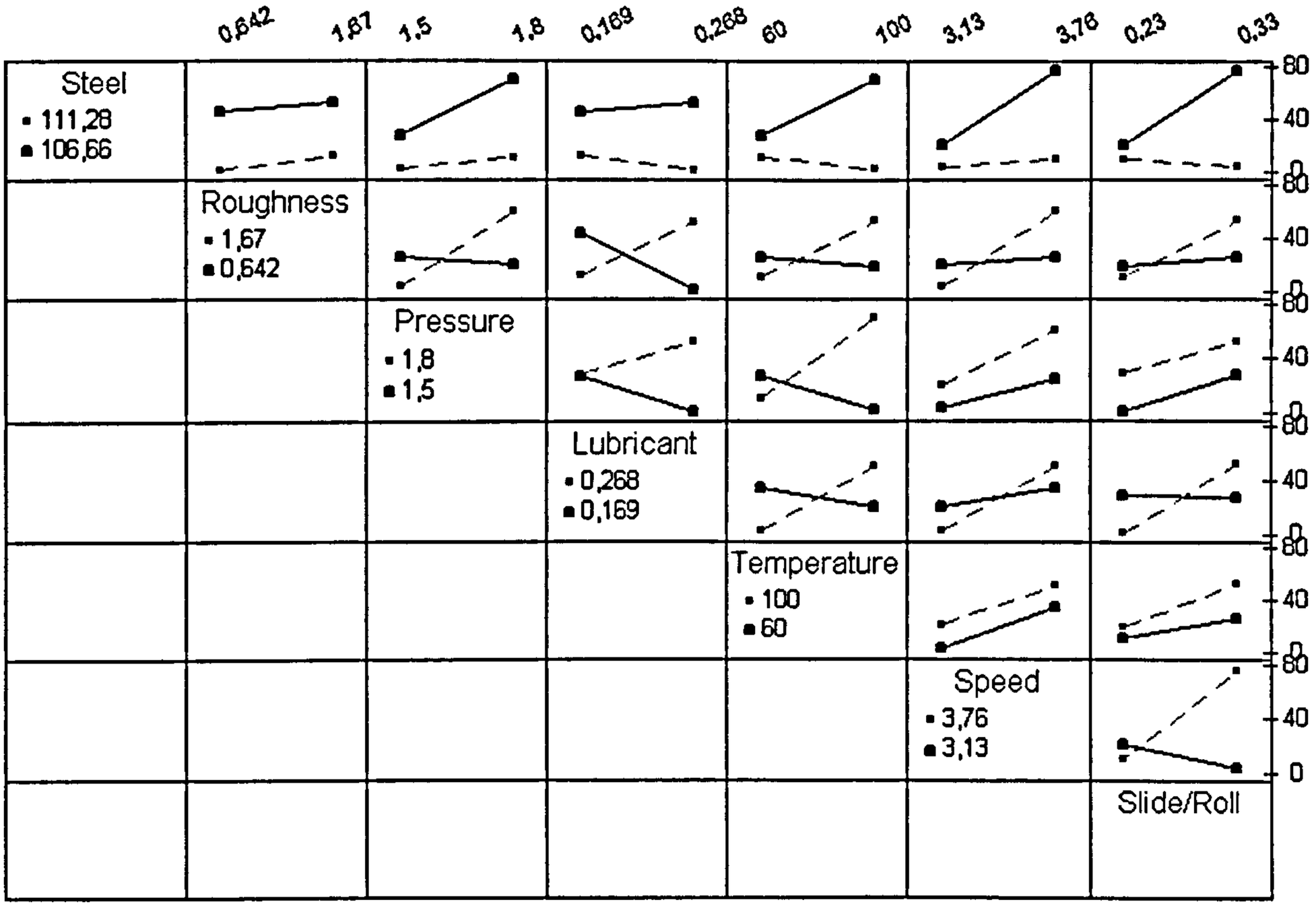


Figure 8-14. The interactions plots for dark etching regions.

Although the effect of temperature is not the biggest, high speed and slide-to-roll ratio generate frictional heating and the temperature is actually higher than the values used in this model. The interaction between temperature and pressure is significant. At high temperature and pressure the *d* factor tends to maximum value.

Table 8-10. The size of effects. Response is *d*.

Factor	Size of effect
Steel	-41,75
Roughness	9,25
Pressure	25,75
Lubricant	-1,25
Temperature	17,25
Speed	30,75
Slide/Roll	24,25

8.4.4.3 White etching bands

The white etching bands have been found in both steels and in gears too but they developed as continuous bands, hence measurable, only in EN36 steel discs. Therefore, the WEB transformation is not specific only to EN36 steel grade as the main effects plots from Figure 8-15 indicates. The significant influences are exerted by temperature, pressure, speed and slide-to-roll ratio, as for the DER factor.

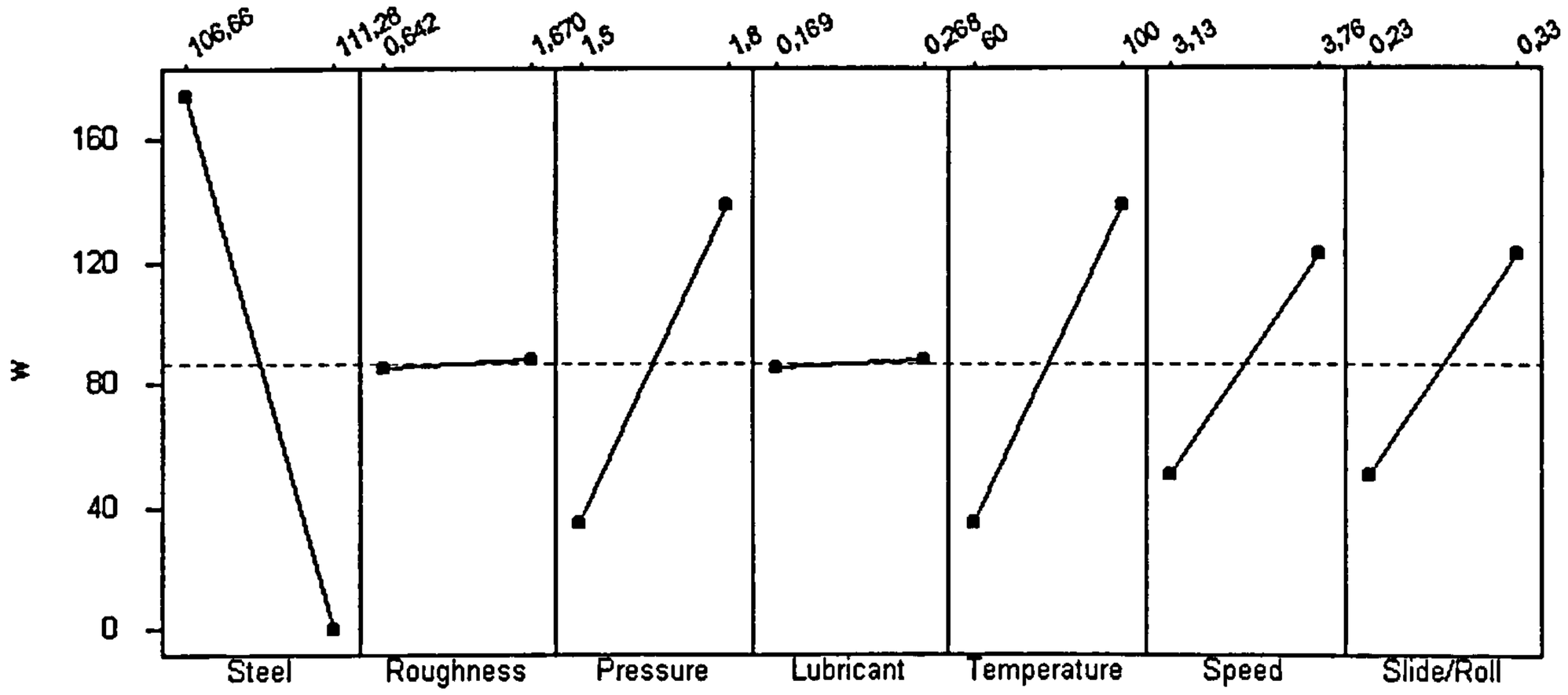


Figure 8-15. The main effects on the development of the white etching bands.

The temperature-pressure interaction for *w* (see Figure 8-16) shows the same characteristic as the interaction for *d* (Figure 8-14). The DER transformation and WEB transformation are favoured by high temperature and contact pressure. This is expected since both phases are products of martensite decay. The effects of speed and slide-to-roll ratio are also significant. The WEB phase is more likely to occur at high speed and slide-to-roll ratio. The interaction between speed and slide-to-roll ratio shows the DER phase does not occur at low speed and slide-to roll ratio.

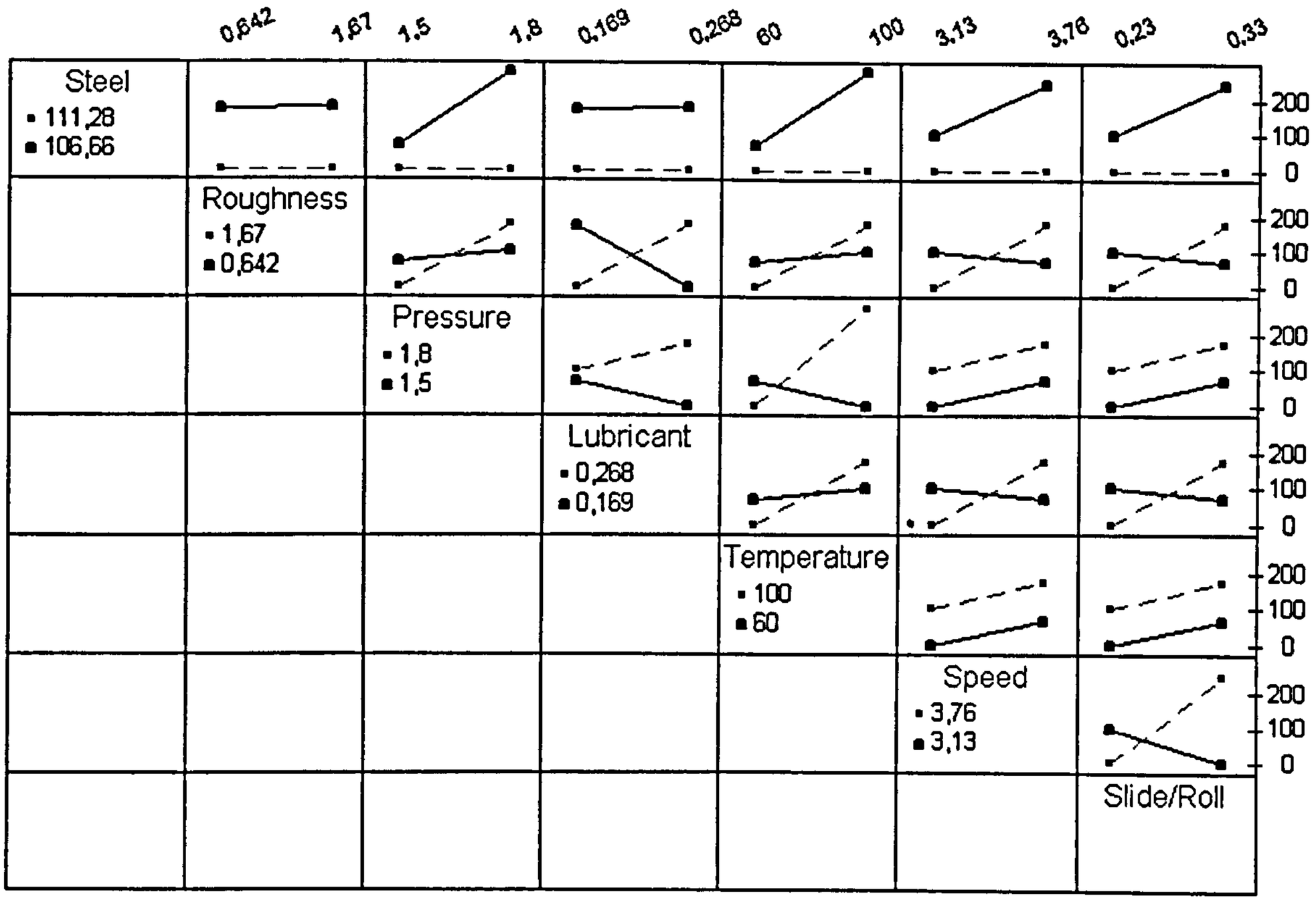


Figure 8-16. The interactions plots for the white etching bands.

Table 8-11. The size of effects. Response is *w*.

Factor	Size of effect
Steel	-173,50
Roughness	2,50
Pressure	103,50
Lubricant	2,50
Temperature	103,50
Speed	72,50
Slide/Roll	72,50

8.5 Summary

In this chapter the influence of seven factors on micropitting has been assessed. The fractional factorial design has been applied twice. At first in order to assess the influence of the factors on micropitting initiation and then in order to analyse the influence of the factors on micropitting propagation.

The initiation stage has been analysed using the number of cycles, N_0, necessary for micropitting to initiate and the lubricant specific film thickness (or lambda ratio), λ, as the responses in the statistical calculations. The propagation of micropitting has been studied using as responses in the statistical calculations the following parameters: the micropitting propagation rate, dM/dN, the specific film thickness, λ, and the phase transformation factors, p, d, and w.

8.5.1 Micropitting initiation

The most significant factor that influences micropitting initiation is the contact pressure. At high contact pressure, the number of cycles necessary for micropitting to initiate is almost independent of the other factors. This implies the existence of a load threshold above which the means to avoid micropitting occurrence are limited. Under the experimental conditions used in this work the only way to retard micropitting is represented by a surface finish improvement. Although both lubricants act better at high temperature (see the interaction lubricant/temperature in Figure 8-2) this effect is reversed by the influence of the other factors resulting in accelerated micropitting occurrence at high temperature.

The biggest influence on the lambda ratio is exerted by temperature. At high temperature the lambda ratio decreases with both types of surface roughness and lubricants. This explains the accelerated micropitting occurrence at high temperature. At high temperature the lubricant film thickness decreases and more asperities collide with each other.

8.5.2 Micropitting propagation

The main effects on micropitting propagation rate, d*M*/d*N*, are those of speed and slide-to-roll ratio. At high speed micropitting propagates with high rate regardless the values of the other factors. This implies that little can be done to stop the progression of micropitting when the operating speed is high. Unlike the initiation, where SAE 8620 steel is more prone to micropitting occurrence, the propagation rate is significantly higher for EN36 steel. As during the initiation period, temperature has an important influence on the lambda ratio during the propagation period. The biggest influences on the crack angle, α, are exerted by the wavelength ratio, Λ and elastic contact modulus, E'.

All the phase transformations strongly depend on the material (steel grade). Material is the biggest factor for all forms of martensite decay. The plastic deformation layer (PDL), the dark etching regions (DER), and the white etching bands (WEB) are more extensively developed in EN36 steel. Considering that the micropitting propagation rate is significantly higher for EN36 steel it can be concluded that micropitting develops faster if martensite decay is more pronounced. This suggests that the mechanism of micropitting and the mechanism of martensite decay are related to each other. Based on this observation it can be assumed that a suitable choice of gear steel, or heat treatment could represent a solution to prevent micropitting.

The second biggest effects on martensite decay are those of temperature, speed and slide-to-roll ratio. Since high speed and slide-to-roll ratio produce high friction, hence high temperature it can be concluded that the driving force for the martensite decay is the temperature.

Further discussion of the results and the conclusions drawn from this work are presented in the next chapter.

~ Chapter 9 ~

DISCUSSION AND CONCLUSIONS

9 DISCUSSION AND CONCLUSIONS

9.1 Summary

The aim of this project was to describe the mechanism of micropitting in gears and to assess the influence of the controlling factors. The main part of the work was carried out on disc specimens tested in conditions similar to those encountered in the gear tooth contact. In order to validate the results obtained from the disc tests, similar investigations were carried out on gears tested at the Design Unit – Gear Technology Centre, Newcastle University. Topographic and metallographic examinations have shown that the two-disc test reliably reproduces micropitting in gears.

The bidirectional distribution of micropits on the affected surface, which can be approximated by the surface profile wavelength in two orthogonal directions: parallel and perpendicular to the surface lay suggests that micropitting occurs due to asperity contact. The surface inside micropits shows the characteristics of a surface resulting from ductile fracture, which has undergone plastic deformation prior to fracture. The depth of maximum shear stress induced by an asperity is in the same range as the depth of the observed cracks which confirms that micropitting is caused by plastic deformation of asperities followed by the initiation and propagation of cracks which ultimately leads to material removal.

The metallographic investigation of the fatigued specimens both, discs and gears, has revealed a number of microstructural features that are not present in untested specimens. Near the surface, zones that exhibit a non-martensitic microstructure were found. It is believed that these are highly deformed regions and the hardness determined by nanoindentation confirms this hypothesis (H = 13.5 GPa). The plastic deformation regions can appear as semi-circular zones, usually less than 10 μm diameter, or as a continuous layer. They were referred to as the plastic deformation layer (PDL).

At higher depths the products of martensite decay were observed. The transformation products are referred to as: dark etching regions (DER), white etching bands (WEB) and butterflies.

It has been found that the boundaries of the plastic deformation regions are preferential sites for microcrack initiation and propagation. If the plastic deformation region has developed as a continuous layer then the cracks initiate and propagate along the layer's boundary, parallel to the surface. If the plastic deformation region has developed as a semi-circular zone beneath an asperity the crack initiates at the surface, at the intersection point between the surface and the boundary and propagates along the boundary, in a direction opposite to the direction of sliding.

The dark etching regions initiate at the prior austenite grain boundary. The DER microstructure is fine and homogeneous as the nanoindentation data indicate. The white etching bands are slightly harder (H = 11.5 GPa) than the original martensite (H=10GPa) but the space between the bands is much harder (H = 20 GPa) due to intense carbide precipitation in these regions. The butterflies observed in this study are not associated with non-metallic inclusions but they are regions of dark etching phase. The butterfly's wings develop symmetrically to the prior austenite grain boundary.

9.2 Micropitting mechanism

It has been found that micropitting develops more extensively in specimens in which the martensite decay is more pronounced. This suggests that the mechanisms by which the two phenomena take place are related to each other. Based on this observation the following micropitting mechanism is proposed.

9.2.1 Micropitting initiation

The values of the specific film thickness, λ, and the friction coefficient, μ, indicate a regime of mixed lubrication in which asperity contact plays a significant role.

When an asperity collides with another one from the mating surface it deforms plastically in the direction of sliding as suggested in Figure 9-1. Due to alternating stresses, plastic deformation extends in the areas beneath asperity and the dislocation density increases. The change in dislocation structure causes work hardening. At the same time, temperature is high enough to activate the diffusion of carbon. As a result, within the boundary of the plastic deformation region (PDR) recrystallisation occurs. The newly formed phase is the dark etching region (DER). The dislocations pile up against this boundary and form a slip band by the intrusion/extrusion mechanism. As cycling continues, dislocations accumulate in the slip band and a crack initiates. Once initiated the crack propagates along the boundary where the dark etching regions exhibits a lower hardness.

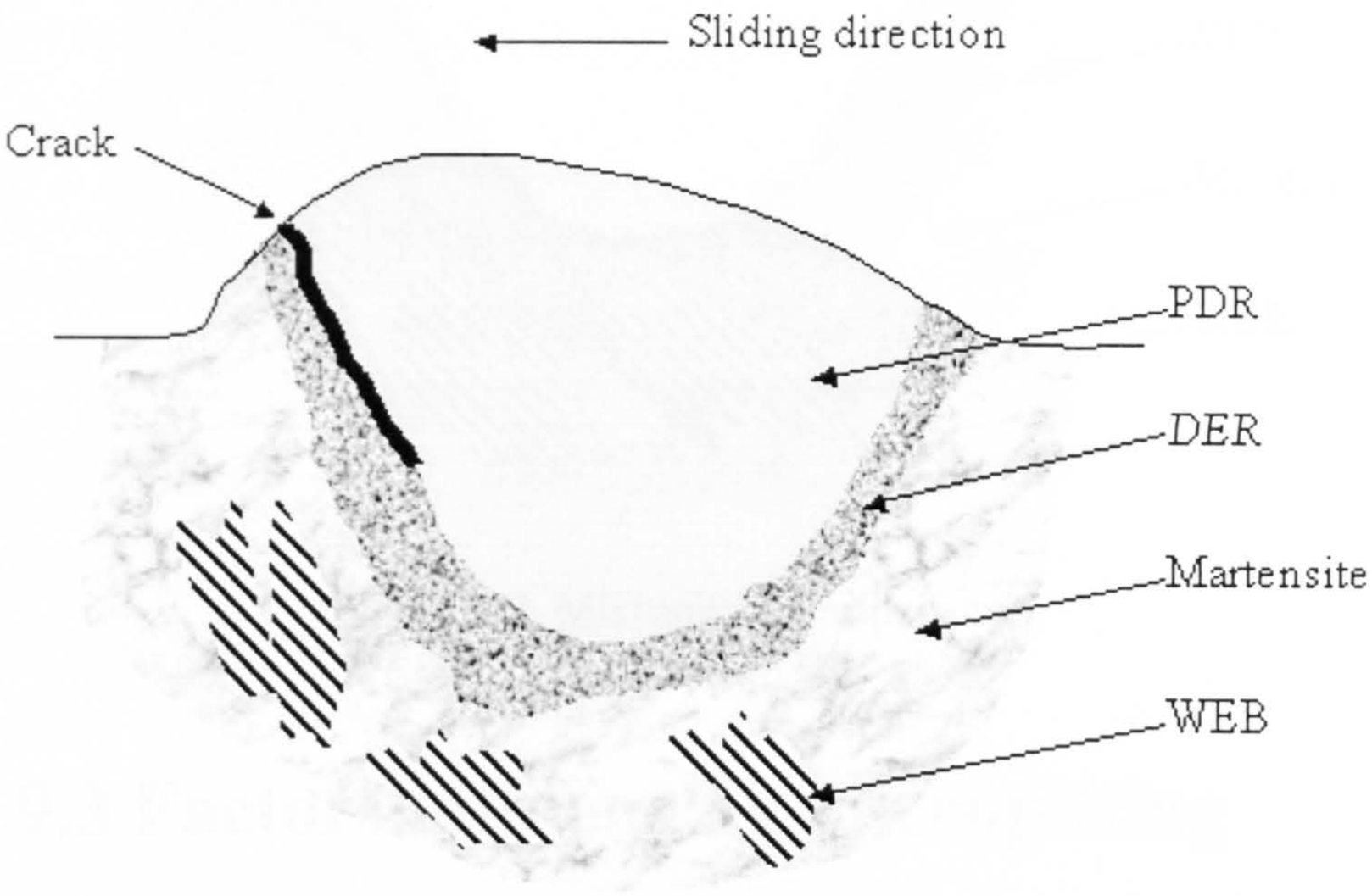


Figure 9-1. Micropitting initiation.

9.2.2 Micropitting propagation

The decay of martensite progresses towards the bulk as cycling continues. Once a micropit has been formed the peaks remaining from the asperity are subjected to the mechanism described above. A crack will initiate at the surface, in the DER phase and starts to propagate (see Figure 9-2). The DER phase that forms at the prior austenite grain boundary represents the preferential site for crack propagation. Additionally, the development of white etching bands provides regions where a crack is more likely to

propagate, i.e., the space between the bands. Therefore crack branching occurs until a complex network of cracks forms below the surface. Every time one of these branches reaches the surface the volume of the material removed from the surface increases hence the micropit widen and deepen. It is also possible that two cracks generated below two neighbouring asperities intersect each other resulting in a wider and deeper micropit.

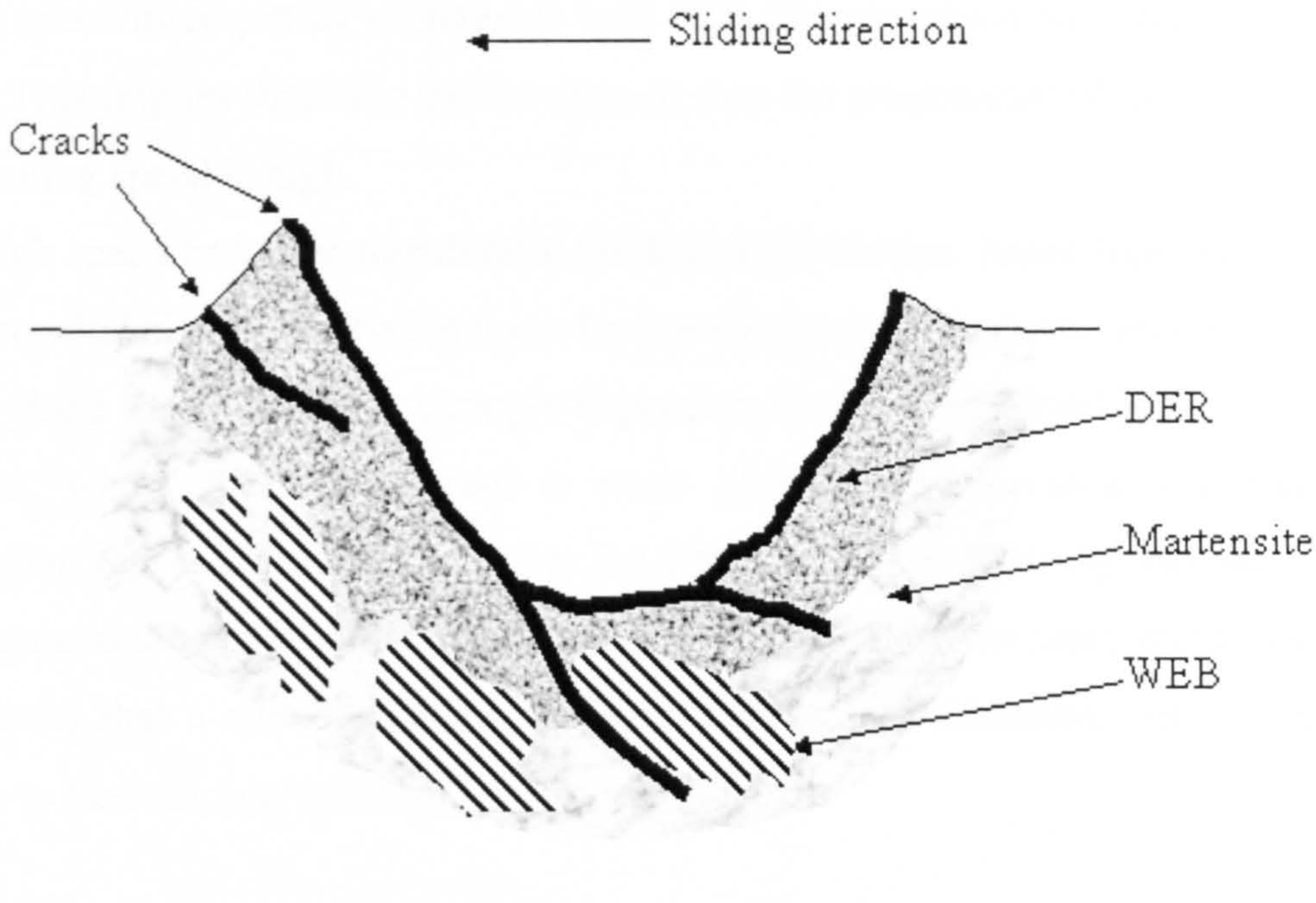


Figure 9-2. Micropitting propagation.

9.3 Factors influencing micropitting

In this work the influence of seven factors has been assessed. The factors are: material, surface lay, load, lubricant, temperature, speed, and slide-to-roll ratio.

9.3.1 Micropitting initiation

The most significant factor that influences micropitting initiation is the contact pressure. At high contact pressure, the number of cycles necessary for micropitting to initiate is almost independent by the other factors. This implies the existence of a load threshold above which the means to avoid micropitting occurrence are limited. Under the

experimental conditions used in this work the only way to retard micropitting is represented by a surface finish improvement.

9.3.2 Micropitting propagation

The main effects on micropitting propagation are those of speed and slide-to-roll ratio. At high speed micropitting propagates with high rate regardless the values of the other factors. This implies that little can be done to stop the progression of micropitting when the operating speed is high.

Since high speed and slide-to-roll ratio produce high friction, hence high temperature, it can be concluded that the driving force for martensite decay is the temperature.

All the phase transformations strongly depend on the material (steel grade). Unlike the initiation, where SAE 8620 steel is more prone to micropitting occurrence, the propagation rate is significantly higher for EN36 steel. Considering that the micropits formation and the phase transformation can be explained by the same mechanism it can be assumed that a suitable choice of gear steel, or heat treatment could represent a solution to prevent micropitting.

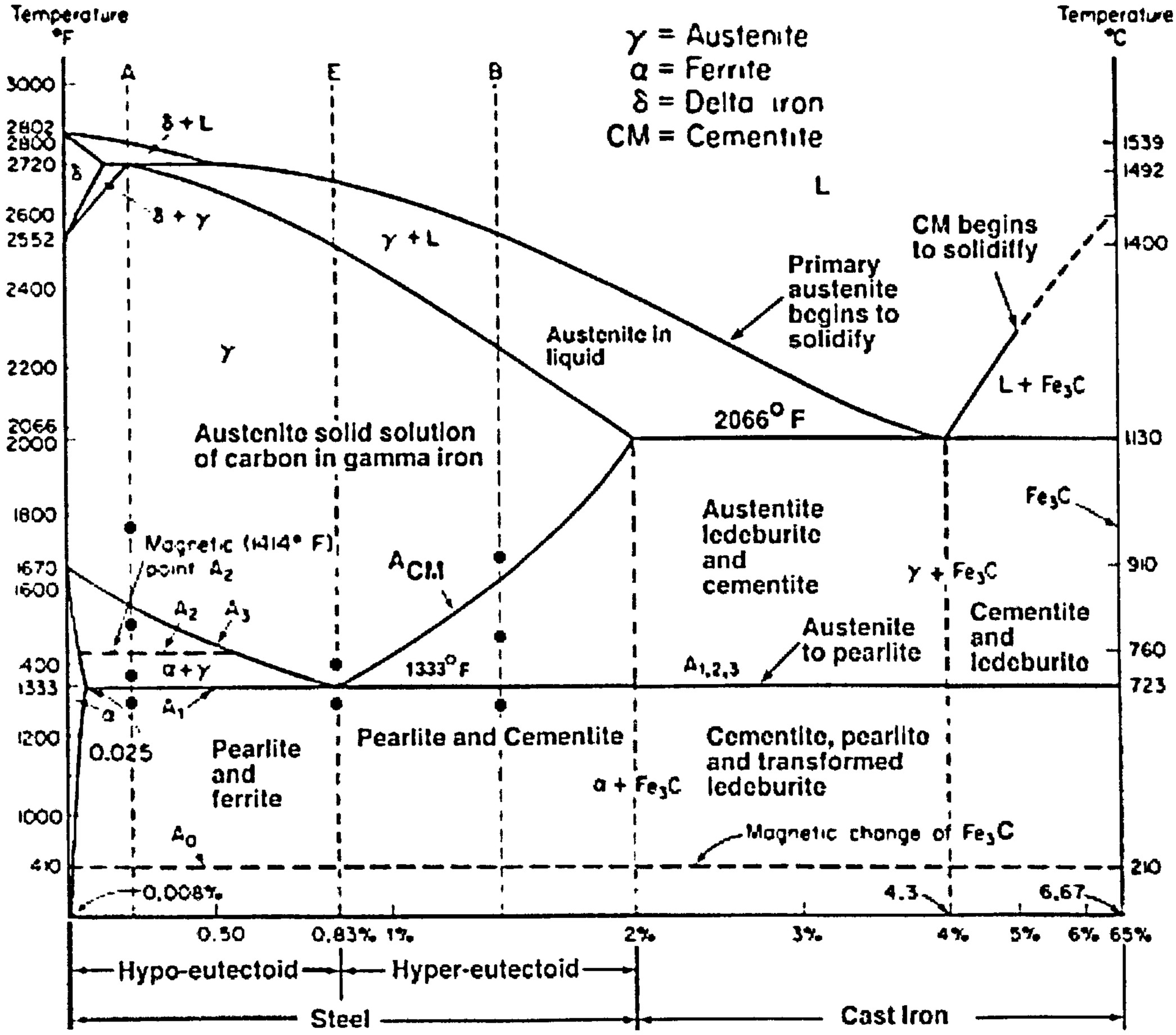
9.4 Further work

To the author's knowledge the current work is the first attempt to explain the micropitting phenomenon through a metallurgical approach. Given the big effect of material on micropitting, further research into the role of microstructure is needed. The size of the prior austenite grain could play a very important role, since the martensite starts to decay at the prior austenite grain boundary.

The most important controlling factor seems to be the temperature. A model that predicts temperature at the asperity level for different contact pairs is necessary. Coatings and surface treatments which reduce the temperature generated during operation are worthy of further research.

APPENDIX 1

Fe-Fe_3C phase diagram



REFERENCES

Abbot, E.J. & Firestone, F.A. (1933) Specifying surface quality. *Mechanical Engineering* 55: 569-572.

Abdel-Al, H.A. (1997) A remark on the flash temperature theory. *International Communications in Heat & Mass Transfer* 24(2): 241-250.

Adams, G.G. & Nosonovsky, M. (2000) Contact modelling - forces. *Tribology International* 33: 431-442.

Amontons, G. (1699) De la resistance causee dans les machines. *Mémoires de l'Académie Royale* A: 257-282.

Archard, J.F. (1958) The temperature of rubbing surfaces. *Wear* 2: 438-455.

Ariura, Y., Ueno, T., Nakanashi, T. (1983) An investigation of surface failure of surface hardened gears by scanning electron microscopy observation. *Wear* 87: 305-316.

ASM Handbook Committee. (1975) Failures of gears. In *Metals Handbook*: *Failure Analysis and Prevention* 10, 8th Edition. Metals Park, Ohio: American Society for Metals.

ASM Committee on Gas Carburizing. (1977) *Carburizing and Carbonitriding*. Metals Park, Ohio: American Society for Metals.

Aylott, C.J. (1998) The effect of lubricant additives on micro-pitting in gears. Stage 3 Materials Project. University of Newcastle.

Bain, E.C. & Dunkirk, N.Y. (1924) The nature of martensite. *Transactions of the Metallurgical Society of AIME* 70: 25-46

Bala, V., Brandt, G., Walters, D.K. (2000) Fuel economy of multigrade gear lubricants. *Industrial Lubrication and Tribology* 52(4): 165-173.

Bartz, W.J. (1993) *Lubrication of gearing*. London: Mechanical Engineering Publications Ltd.

Bartz, W.J. (2000) Gear oil influences on efficiency of gear and fuel economy of cars. *Proceedings of the Institution of Mechanical Engineers* 214(D): 189-196.

Barus, C. (1893) Isotherms, isopiestics and isometrics relative to viscosity. *American Journal of Science* 45: 87-96.

Batista, A.C., Dias, A.M., Lebrun, J.L., Le Flour, J.C., Inglebert, G. (2000) Contact fatigue of automotive gears: evolution and effects of residual stresses introduced by surface treatments. *Fatigue and Fracture of Engineering Materials and Structures* 23: 217-228.

Berthe, D., Michau, B., Flamand, L., Godet, M. (1978) Effect of roughness ratio and Hertz pressure on micro-pits and spalls in concentrated contacts: theory and experiments. In D. Dowson (ed.) *Surface roughness effects in lubrication: Proceedings of 4th Leeds-Lyon Symposium on Tribology*, London-Lyon: Mechanical Engineering Publications and Institut National des Sciences Appliqués de Lyon.

Berthe, D., Flamand, L., Foucher, D., Godet M. (1980) Micropitting in Hertzian contacts. *Transactions of the ASME - Journal of Lubrication Technology* 102: 478-489.

Bhargava, V., Hahn, G.T., Rubin, C.A. (1990) Rolling contact deformation, etching effects, and failure of high-strength bearing steel. *Metallurgical Transactions* A 21A: 1921-1931.

Bhushan, B. (2001) *Modern Tribology Handbook*. Boca Raton, Fl: CRC Press.

Blok, H. (1937) Theoretical study of temperature rise at surface of actual contact under oilness lubricating conditions. *Proceedings of the Institution of Mechanical Engineers* 2: 222-235.

Bower, A.F. (1988) The influence of crack face friction and trapped fluid on surface initiated rolling contact fatigue cracks. *Transactions of the ASME-Journal of Tribology* 110: 704-711.

Brechot, P., Cardis, A.B., Murphy, W.R., Theissen, J. (2000) Micropitting resistant industrial gear oils with balanced performance. *Industrial Lubrication and Tribology* 52(3): 125-136.

Britton, R.D., Elcoate, C.D., Alanou, M.P., Evans, H.P., Snidle, R.W. (2000) Effect of surface finish on gear tooth friction. *Transactions of the ASME-Journal of Tribology* 122: 354-360.

Brookfield Engineering Laboratories, Inc. (2000) More solutions to sticky problems. Technical paper on viscosity measurement and control. Middleboro, MA. Retrieved May 16, 2000 from World Wide Web: http://www.brookfieldengineering.com.

BS 436: 1986 Part 3: *Strength and Load Carrying Capacity of Gears*. London: British Standard Institution.

BS 6741: 1987 *British Standard Glossary of Surface roughness terms. Part 1. Surface and its parameters*. London: British Standard Institution.

BS 7848: 1996 *Gears - Wear and damage to gear teeth - Terminology*. London: British Standard Institution.

BS 14635-1: 2000 Gears - FZG test procedures - Part 1: FZG test method A/8, 3/90 for relative scuffing load-carrying capacity of oils. London: British Standard Institution.

Buchwald, J. & Heckel, R.W. (1968) An analysis of microstructural changes in 52100 steel bearings during cyclic stressing. *Transactions of the ASM* 61: 750-756.

Bull, S.J., Evans, J.T., Korsunsky, A.M., Page, T.F. (1997) Can novel vapour deposited coatings prevent micropitting in gears? Draft Case for Support for EPSRC proposal to support GRF micropitting work. University of Newcastle.

Bull, S.J., Evans, J.T., Shaw, B.A., Hofmann, D.A. (1999) The effect of white layer on micro-pitting and surface contact fatigue failure of nitrided gears. *Proceedings of the Institution of Mechanical Engineers* 213(J): 305-313.

Bull, S.J. (2002) Extracting hardness and Young's modulus from nanoindentation load-displacement curves. *Zeitschrift für Metallkunde* 93: 1-5.

Bush, J.J., Grube, W.L., Robinson, G.H. (1961) Microstructural and residual stress changes in hardened steel due to rolling contact. *Transactions of the ASM* 54: 390-412.

Cameron, A. (1981) *Basic Lubrication Theory* (3rd edition). Chichester: Ellis Horwood Ltd.

Cardis, A.B. & Webster, M.N. (2000) Gear oil micropitting evaluation. *Gear Technology* 17 (5): 30-35.

Castro, J. & Seabra, J. (1998) Scuffing and lubricant film breakdown in FZG gears. Part I: Analytical and experimental approach. *Wear* 215: 104-113.

Cavallaro, G.P., Wilks, T.P., Subramanian, C., Strafford, K.N., French, P., Allison, J.E. (1995) Bending fatigue and contact fatigue characteristics of carburized gears. *Surface and Coatings Technology* 71:182-192.

Chang, L. (1995) A deterministic model for line-contact partial elasto-hydrodynamic lubrication. *Tribology International* 28 (2): 75-84.

Chang, L. & Zhao, Y. (2000) On the sensitivity of the asperity pressures and temperatures to the fluid pressure distribution in mixed-film lubrication. *Transactions of the ASME - Journal of Tribology* 122: 77-85.

Cheng, Y.-T. & Cheng, C.-M. (2000) What is indentation hardness? *Surface & Coatings Technology* 133-134: 417-424.

Choi, B.Y. & Bahng, G.W. (1985) Characterisation of microstructure and its effect on rolling contact fatigue of induction hardened medium carbon bearing steels. *Materials Science and Technology* 14: 816-821.

Chue, C.H., Chung, H.H., Lin, J.F., Chou C.C. (2001) The effects of strain hardened layer on pitting formation during rolling contact. *Wear* 249: 109-116.

Cochran, W.G. & Cox, G.M. (1992) *Experimental designs* (2nd edition). New York: Wiley.

Cottrell A.H. & Hull D. (1957) Extrusion and intrusion by cyclic slip in copper. *Proceedings of the Royal Society* 242A: 211-213.

Cram, W.D. (1961) Experimental load-stress factors. In C. Lipson & L.V. Colwell (eds.) *Handbook of Mechanical Wear. Wear, Frettage, Pitting, Cavitation, Corrosion.* Ann Arbor: The University of Michigan Press.

Dabnichki, P. & Crocombe, A. (1999) Finite element modelling of local contact conditions in gear teeth. *Journal of Strain Analysis* 34(2): 129-142.

Dawson, P.H. (1962) Effect of metallic contact on the pitting of lubricated rolling surfaces. *Journal of Mechanical Engineering Science* 4(1): 16-21.

DeLange, G. (2000) Failure Analysis for Gearing. *Maintenance Technology*. Applied Technology Publications, Inc. Retrieved May 16, 2000 from World Wide Web: http://www.mt-online.com.

Dieter, G.E. (1988) *Mechanical Metallurgy*. New York: McGraw-Hill.

Dowson, D. & Higginson, G.R. (1959) A numerical solution to the elastohydrodynamic problem. *Journal of Mechanical Engineering Science* 1(1): 6-15.

Dowson, D. & Higginson, G. R. (1977) *Elasto-hydrodynamic Lubrication*, Oxford: Pergamon Press Ltd.

Dowson, D. (1979) *History of Tribology* London: Longman Group Ltd.

Dowson, D. & Ehret, P. (1999) Past, present and future studies in elastohydrodynamics. *Proceedings of the Institution of Mechanical Engineers* 213(J): 317-333.

Dölle, H. & Cohen, J.B. (1980) Residual stresses in ground steels. *Metallurgical Transactions* A (11A): 159-164.

Dudley D.W. (1962) *Gear Handbook. The Design, Manufacture, and Application of Gears*. New York: McGraw-Hill.

Errichello R. (2003) Selecting and Applying Lubricants to Avoid Micropitting of Gear Teeth. *Machinery Lubrication Magazine,* 200211, November. Retrieved January 03, 2003 from World Wide Web: http://www.machinerylubrication.com.

Ericsson, T. & Hildenwall, B. (1982) Thermal and transformation stresses. In E. Kula and V. Weiss (eds.) *Residual Stress and Stress Relaxation*, Army Materials Research Conference, New York: Plenum Press.

Evans, H.P. & Snidle, R.W. (1996) Analysis of micro-elastohydrodynamic lubrication for engineering contacts. *Tribology International* 29(8): 659-667.

Ewing J.A. & Humphrey J.C.W. (1903) *Philosophical Transactions of the Royal Society* 200: 241.

Fernandes, P.J.L. & McDulling, C. (1997) Surface contact fatigue failures in gears. *Engineering Failure Analysis* 4(2): 99-107.

Fine, M.E. (1980) Fatigue resistance of metals. *Metallurgical Transactions* A 11A: 365-379.

Gao, J., Lee, Si.C., Ai, X., Nixon, H. (2000) An FFT-based transient flash temperature model for general three-dimensional rough surface contacts. *Transactions of the ASME - Journal of Tribology* 122: 519-523.

Gentile, A.J., Jordan, E.F., Martin, A.D. (1965) Phase transformations in high-carbon, high-hardness steels under contact loads. *Transactions of the Metallurgical Society of AIME* 233: 1085-1093.

Glodez, S., Flasker, J., Ren, Z. (1997) A new model for the numerical determination of pitting resistance of gear teeth flanks. *Fatigue and Fracture of Engineering Materials and Structures* 20(1): 71-83.

Graham, J.D. (1961) Pitting of gear teeth. In C. Lipson and L.V. Colwell (eds.) *Handbook of Mechanical Wear. Wear, Frettage, Pitting, Cavitation, Corrosion*. Ann Arbor: The University of Michigan Press.

Graham, R.C., Olver, A., Macpherson, P.B. (1981) An Investigation into the mechanism of pitting in high-hardness carburized steels. *ASME Conference, Century 2 International Power Transmissions & Gearing*, San Francisco, 18-21 August, 1980, 80-C27DET-118.

Greenwood, J.A. & Williamson, J.B.P. (1966) Contact of nominally flat surfaces. *Proceedings of the Royal Society* A295(1442): 300-319.

Greenwood, J.A. (1992) Contact of rough surfaces. In I.L. Singer & H.M. Pollock (eds.) *Fundamentals of Friction: Macroscopic and Microscopic Processes*. Dordrecht: Kluwer Academic Publishers.

Greenwood, J.A. (1999) Transverse roughness in elastohydrodynamic lubrication. *Proceedings of the Institution of Mechanical Engineers* 213(J): 383-396.

Griffith, A.A. (1920) The phenomena of rupture and flow in solids. *Philosophical Transactions of The Royal Society*, 221A: 163-198.

Grubin, A.N. (1949) Fundamentals of the hydrodynamic theory of lubrication of heavily loaded cylindrical surfaces. In K.F. Ketova (ed.) *Investigation of the Contact of Machine Components, Central Scientific Research Institute for Technology and Mechanical Engineering*, Moscow, Book 30, (DSIR Translation 337): 115-166.

Guagliano, M., Riva, E., Guidetti, M. (2002) Contact fatigue failure analysis of shot-peened gears. *Engineering Failure Analysis* 9: 147-158.

Guangteng, G., Cann, P.M., Olver, A.V., Spikes, H.A. (2000) Lubricant film thickness in rough surface, mixed elastohydrodynamic contact. *Transactions of the ASME-Journal of Tribology* 122: 65-76.

Guha, D. & Roy Chowdhuri, S.K. (1996) The effect of surface roughness on the temperature at the contact between sliding bodies. *Wear* 197: 63-73.

Gunther, R. C. (1972) *Lubrication*. Folkestone: Bailey Brothers & Swinfen Ltd.

Hall, E.O. (1951) The deformation and ageing of mild steel: III Discussion of results. *Proc. Phys. Soc.* B64: 747-753.

Hammersley, G., Hackel, L.A., Harris, F. (2000) Surface prestressing to improve fatigue strength of components by laser shot peening. Paper presented at *Materials Congress 2000*, Cirencester, UK.

Hamrock, B.J. & Dowson, D. (1976a) Isothermal elastohydrodynamic lubrication of point contacts. Part I-theoretical formulation. *Transactions of the ASME - Journal of Tribology* 98(2): 223-229.

Hamrock, B.J. & Dowson, D. (1976b) Isothermal elastohydrodynamic lubrication of point contacts. Part II-ellipticity parameter results. *Transactions of the ASME - Journal of Tribology* 98(3): 375-383.

Hamrock, B.J. & Dowson, D. (1977) Isothermal elastohydrodynamic lubrication of point contacts. Part III-fully flooded results. *Transactions of the ASME - Journal of Tribology* 99(2): 264-276.

Hay, J.C., Bolhakov, A., Pharr, G.M. (1999) A critical examination of the fundamental relations used in the analysis of nanoindentation data. *Journal of Materials Research* 14(6): 2296-2305.

Hertz, H. (1881) Ueber die Berührung fester elasticher Körper. *Journal für die Reine und Angewandte Mathematik* 92: 156-171.

Hoeprich, M.R. (2000) Analysis of micro-pitting on prototype surface fatigue test gears. *12th International Colloquium Tribology 2000-Plus*, 12-13 January, Technische Akademic Esslingen, 3: 2023-2033.

Hofmann, D.A., Kohler H.K. & Munro R.G. (1991) *Gear Technology*. Burton on Trent: British Gear Association.

Hollande, S. (1987) Understanding micro-pitting in gears. MSc thesis. University of Newcastle.

Holmberg, K. & Matthews, A. (1994) *Coatings Tribology. Properties, Techniques and Applications in Surface Engineering*. Amsterdam: Elsevier.

Honeycombe, R.W.K. (1981) *Steels: Microstructure and Properties*, London: Edward Arnold.

Höglund, E. (1998) Influence of lubricant properties on elastohydrodynamic lubrication. *Wear* 232: 176-184.

Höhn, B.-R., Oster, P., Emmert, S. (1996) Micropitting in case-carburized gears - FZG micropitting test. *VDI Berichte* 1230: 331-344.

Ioannides, E. & Kuijpers, J.C. (1986) Elastic stresses below asperities in lubricated contacts. *Transactions of the ASME - Journal of Tribology* 108: 394-402.

Jacobson, Bo (2000) Thin film lubrication of real surfaces. *Tribology International* 33: 205-210.

Jaegar, J.C. (1942) Moving sources of heat and the temperature at sliding surfaces. *Proceedings of the Royal Society* 66: 203-204.

Johnson, K.L. (1987) *Contact Mechanics*. Cambridge: Cambridge University Press.

Jones, A.B. (1946) Metallographic observations of ball bearing fatigue phenomena. *Proceedings of the American Society for Testing Materials (ASTM)* 46: 1-6

Kaneta, M. and Nishikawa, H. (1999) Experimental study on microelastohydrodynamic lubrication. *Proc. Instn. Mech. Engrs.* 213(J): 371-381.

Kandil, F.A., Lord, J.D., Fry, A.T., Grant, P.V. (2001) A review of Residual Stress Measurement Methods - A Guide to Technique Selection. NPL Materials Centre: NPL Report MATC(A)04.

Kapoor, A., Johnson, K.L. (1994) Plastic ratchetting as a mechanism of metallic wear. *Proceedings of the Royal Society* A 445: 367-381.

Krauss, G. (1995a) Microstructure and performance of carburized steel, Part I: Martensite. *Advanced Materials & Processes* 5: 40Y-40BB.

Krauss, G. (1995b) Microstructure and performance of carburized steel, Part II: Austenite. *Advanced Materials & Processes* 7: 48U-48Y.

Krauss, G. (1995c) Microstructure and performance of carburized steel, Part III: Austenite and fatigue. *Advanced Materials & Processes* 9: 42EE-42II.

Krauss, G. (1999) Martensite in steel: strength and structure. *Materials Science and Engineering* A(273-275): 40-57.

Laird, C. (1979) Mechanisms and theories of fatigue. In *Fatigue and Microstructure.* Papers presented at the 1978 ASM Materials Science Seminar, 14-15 October, 1978, St. Louis, Missouri.

Lamagnere, P., Girodin, D., Meynaud, P., Vergne, F., Vincent, A. (1996) Study of elasto-plastic properties of microheterogeneities by means of nano-indentation measurements: application to bearing steels. *Materials Science & Engineering* A215: 134-142.

Lansdown, A.R. (1996) *Lubrication and Lubricant Selection. A Practical Guide.* London: Mechanical Engineering Publications.

Larsson, R., Larsson, P.O., Eriksson, E., Sjöberg, M., Höglund, E. (2000) Lubricant properties for input to hydrodynamic and elastohydrodynamic lubrication analyses. *Proceedings of the Institution of Mechanical Engineers* 214(J): 17-27.

Leslie, W.C. (1981) *The Physical Metallurgy of Steels*. New York: McGraw-Hill.

Littman, W.E. & Widner, R.L. (1966) Propagation of contact fatigue from surface and subsurface origins. *Transactions of the ASME - Journal of Basic Engineering*, September: 624-636.

Lo, C.C. (1969) Elastic contact of rough cylinders. *International Journal of Mechanical Science* 11: 105-115.

Lubrecht, A.A. & Venner, C.H. (1999) Elastohydrodynamic lubrication of rough surfaces. *Proceedings of the Institution of Mechanical Engineers* 213(J): 397-404.

MatWeb.com (2003) The Online Materials Database. Automation Creations, Inc. Retrieved January 14, 2003 from World Wide Web: http://www.matweb.com.

Martin, H.M. (1916) Lubrication of gear teeth. *Engineering* 102: 119-121.

Martin, J.A., Borgese, S.F., Eberhardt, A.D. (1966) Microstructural alterations of rolling-bearing steel undergoing cyclic stressing. *Transactions of the ASME - Journal of Basic Engineering*, September: 555-567.

McCool, J.I. (1986) Extending the capability of the Greenwood Williamson microcontact model. *Transactions of the ASME - Journal of Tribology* 122: 496-502.

McGurk, M. (1997) An exploration of contact deformation mechanisms in thin hard coated systems. PhD thesis. University of Newcastle.

Mihailidis, A., Retzepis, J., Salpistis, C, Panajiotidis, K. (1999) Calculation of friction coefficient and temperature field of line contacts lubricated with a non-Newtonian fluid. *Wear* 232: 213-220.

Mikic, B.B. (1974) Thermal contact conductance; theoretical considerations. *International Journal of Heat Mass Transfer* 17: 205-214.

Moore, A.J. (1997) The behaviour of lubricants in elastohydrodynamic contacts. *Proceedings of the Institution of Mechanical Engineers* 211(J): 91-106.

Mott, NF (1958) A theory of the origin of fatigue cracks. *Acta Metallurgica* 6: 195-197.

Mura, T., & Nakasone, Y. (1990) A theory of fatigue crack initiation in solids. *Transactions of the ASME - Journal of Applied Mechanics* 57: 1-6.

Muroga, A. & Saka, H. (1995) Analysis of rolling contact fatigued microstructure using focused ion beam sputtering and transmission electron microscopy observation. *Scripta Metallurgica et Materialia* 33(1): 151-156.

Nakajima, A., Mawatari, T., Yoshida, M., Tani, K., Nakahira, A. (2000) Effects of coating thickness and slip ratio on durability of thermally sprayed WC cermet coating in rolling/sliding contact. *Wear* 241: 166-173.

Nélias, D., Dumont, M.L., Champiot, F., Vincent, A., Girodin, D., Fougeres, R., Flamand, L. (1999) Role of inclusions, surface roughness and operating conditions on rolling contact fatigue. *Transactions of the ASME - Journal of Tribology* 121: 240-251.

Nélias, D. & Ville, F. (2000) Detrimental effects of debris dents on rolling contact fatigue. *Transactions of the ASME - Journal of Tribology* 122: 55-64.

Neumann, P. (1969) Coarse slip model of fatigue. *Acta Metallurgica* 17(9): 1219-1225.

Nikas, G. K., Sayles, R., S., Ioannides, E. (1998) Effect of debris particles in sliding/rolling elastohydrodynamic contacts. *Proceedings of the Institution of Mechanical Engineers* 212(J): 333-343.

Nishiyama, Z. (1978) *Martensite Transformation*. New York: Academic Press.

Nixon, H.P. (1998) Effects of extreme pressure additives in lubricants on bearing fatigue life. *Iron and Steel Engineer* December: 21-26.

Noyan, I.C. & Cohen, J.B. (1982) The nature of residual stress and its measurement. In E. Kula and V. Weiss (eds.) *Residual stress and Stress Relaxation*, New York: Plenum Press.

O'Brien, J.L. & King, A.H. (1966a) Cyclic-stress-induced microstructural changes in hardened steel bearings. *Journal of the Iron & Steel Institute*, January: 55

O'Brien, J.L. & King, A.H. (1966b) Electron microscopy of stress-induced structural alterations near inclusions in bearing steels. *Transactions of the ASME - Journal of Basic Engineering*, September: 568-588.

Oila, A. & Bull, S.J. (2003) Nanoindentation testing of gear steels. *Zeitschrift für Metallkunde* 94 (7): 793-797.

Oliver, W.C. & Pharr, G.M. (1992) An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *Journal of Materials Research* 7(6): 1564-1583.

Oliver, W.C. (2001) An alternative technique for analyzing instrumented indentation data. *Journal of Materials Research* 16: 3202-3211.

Olver, A.V. (1983) Micropitting and asperity deformation. In *Developments in Numerical and Experimental Methods Applied to Tribology. Proceedings of the 10th Leeds-Lyon Symposium on Tribology*, Lyon 6-9 September 1983.

Olver, A.V. (1986) Wear in rolling contacts. *Wear* 112: 121-144.

Olver A.V. (1995) Micro-pitting of gear teeth: design solutions. *IMechE Congress Aerotech '95*, October 1995, Birmingham.

Olver, A.V. & Spikes, H.A. (1998) Prediction of traction in elastohydrodynamic lubrication. *Proceedings of the Institution of Mechanical Engineers* 212(J): 321-332.

Österlund, R. & Vingsbo, O. (1980) Phase changes in fatigued ball bearings. *Metallurgical Transactions* A (11A): 701-707.

Österlund, R., Vingsbo, O., Vincent, L., Guiraldenq, P. (1982) Butterflies in fatigued ball bearings - formation mechanisms and structure. *Scandinavian Journal of Metallurgy* 11: 23-32.

Pandey, R.K. & Ghosh, M.K. (1998) Temperature rise due to sliding in rolling/sliding elastohydrodynamic lubrication line contacts: an efficient numerical analysis for contact zone temperatures. *Tribology International* 31(12): 745-752.

Petch, N.J. (1953) The cleavage strength of polycrystals. *Journal of Iron & Steel Institute* 174: 25-28.

Petrusevich, A.I. (1951) Fundamental conclusions from the contact hydrodynamic theory of lubrication. *Izv. Akad., Nauk. SSSR (OTN)* 2: 209-223.

Pharr, G.M., & Oliver, W.C. Measurements of thin film mechanical properties using nanoindentation. *Materials Research Society Bulletin* 17: 28-33.

Polonsky, I.A. (1995) Scale effects of elastic-plastic behavior of microscopic asperity contacts. *ASME/STLE Tribology Conference*, 8-11 October 1995, Orlando, 95-TRIB-27.

Polonsky, I.A., Chang, T.P., Keer, L.M., Sproul W.D. (1997) An analysis of the effect of hard coatings on near-surface rolling contact fatigue initiation induced by surface roughness. *Wear* 208: 204-219.

Porter, D.A. and Easterling, K.E. (1992) *Phase transformations in metals and alloys*. London-New York: Chapman & Hall.

Prevey, P.S. (1996) Current applications of X-ray diffraction residual stress measurement. In G. Vander Voort & J. Friel (eds.) *Developments in Materials Characterization Technologies*. Materials Park, Ohio: ASM International.

Priestner, R., Priestner, D.M. (1991) The importance of the substrate in surface engineering. In P.H. Morton (ed.) *Surface Engineering & Heat Treatment. Past, present and future*, London: The Institute of Metals.

Rakhit, A.K. (2000) *Heat Treatment of Gears*. Metals Park, Ohio: ASM International.

Reynolds, O. (1886) On the theory of lubrication and its application to Mr. Beauchamp Tower's experiments, including an experimental determination of the viscosity of olive oil. *Philosophical Transactions of The Royal Society* 117: 157-234.

Ringsberg, J.W. (2001) Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue* 23: 575-586.

Roelands, C. J. A. (1966) Correlational Aspects of the Viscosity-Temperature-Pressure Relationship of Lubricating Oils. PhD thesis. Technical University Delft, The Netherlands.

Roylance, B.J., Williams, J.A., Dwyer-Joyce, R. (2000) Wear debris and associated wear phenomena - fundamental research and practice. *Proceedings of the Institution of Mechanical Engineers* 214(J): 79-105.

Schrader, A. & Rose, A. (1966) *De Ferri Metallographia II. Structure of Steels*. Düsseldorf: Verlag Stahleisen m.b.H.

Seireg, A. (2001) Thermal stress effects on the surface durability of gear teeth. *Proceedings of the Institution of Mechanical Engineers* 215(C): 973-979.

Sharif, K.J., Barragan de Ling, FdM., Martin, M.J., Alanou, M.P., Evans, H.P., Snidle R.W. (2000) Film thickness predictions for elastohydrodynamic elliptical contacts over a wide range of radius ratios with consideration of side starvation. *Proceedings of the Institution of Mechanical Engineers* 214(J): 63-78.

Shipley, E.E. (1974) Failure modes in gears. In P.J. Guichelaar, B.S. Levy, N.M. Parikh (eds.) *Gear manufacture and performance*. American Society for Metals.

Shotter, B.A. (1980) Micropitting: its characteristics and implications on the test requirements of gear oils. In R. Tourret & E.P. Wright (Eds.) *Performance and testing of gear oils and transmission fluids - Proceedings of the International Symposium*. London: The Institute of Petroleum.

Smallman, R.E. (1962) *Modern Physical Metallurgy*. London: Butterworths.

Sneddon, I.N. (1965) The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *International Journal for Engineering Science* 3: 47-57.

Snidle, R.W., Evans, H.P., Alanou, M.P. (2000) Gears: elastohydrodynamic lubrication and durability. *Proceedings of the Institution of Mechanical Engineers* 214(C): 39-50.

Steven W. & Haynes A.G. (1956) J. Iron Steel Inst. 183: 349.

Straffelini, G., Colombo, D., Molinari, A. (1999) Surface durability of electroless Ni-P composite deposits, *Wear* 236: 179-188.

Students, E. & Rudzitis, J. (1996) Contact of surface asperities in wear. *Tribology International* 29(4): 275-279.

Styri, H. (1951) Fatigue Strength of Ball Bearing Races and Heat-Treated 52100 Steel Specimens. *Proceedings of the ASTM* 51: 682-700.

Swahn, H., Becker, P.C., Vingsbo, O. (1976) Martensite decay during rolling contact fatigue in ball bearings. *Metallurgical Transactions* A, 7A: 1099-1110.

Szpunar, E. & Bielanik, J. (1984) Influence of retained austenite on propagation of fatigue cracks in carburised cases of toothed elements. In Conference Proceedings *Heat Treatment 84*, London: The Metals Society.

Tallian, T.E. (1992) Simplified contact fatigue life prediction model-Part I: Review of published models. *Transactions of the ASME-Journal of Tribology* 114: 207-213.

Thelning, K.-E. (1975) *Steel and its Heat Treatment-Bofors Handbook*, London & Boston: Butterworths.

Thomson, N.J., Wadsorth, N., Louat, N. (1956) The origin of fatigue fracture in copper. *Philosophical Magazine* 1: 113-126.

Townsend, D.P. (1992) Improvement in surface fatigue life of hardened gears by high-intensity shot peening. *NASA-TM-105678, Technical Report 91-C-042.*

Vander Voort, G.F. (1984) *Metallography, principles and practice.* New York: McGraw-Hill.

Vander Voort, G.F. (2000) Trends in Specimen Preparation. *Advanced Materials & Processes* 157(2): 45-49.

Ville, F. (1998) Pollution solide des lubrifiants, indentation et fatigue des surfaces. PhD thesis. Institut National des Sciences Appliquees de Lyon

Voskamp, A.P., Österlund, R., Becker, P.C., Vingsbo, O. (1980) Gradual changes in residual stress and microstructure during contact fatigue in ball bearings. *Metals Technology*, January: 14-21.

Voskamp, A.P. (1985) Material response to rolling contact loading. *Transactions of the ASME - Journal of Tribology* 107: 359-366.

Way, S. (1935) Pitting due to rolling contact. *Transactions of the ASME-Journal of Applied Mechanics* 2: A49-A58.

Webster, M.N. & Norbart, C.J.J. (1995) An Experimental Investigation of Micropitting Using a Roller Disk Machine. *Tribology Transactions* 38(4): 883-893.

Whitehouse, D.J. (1994) *Handbook of Surface Metrology*. Bristol: Institute of Physics Publishing.

Widmark, M. & Melander, A. (1999) Effect of material, heat treatment, grinding and shot peening on contact fatigue life of carburised steels. *International Journal of Fatigue* 21: 309-327.

Winter, H. & Weiss, T. (1981) Some factors influencing the pitting, micro-pitting (frosted areas) and slow speed wear of surface hardened gears. *Transactions of the ASME - Journal of Mechanical Design* 103: 499-505.

Winter, H. & Oster, P. (1990) Influence of lubrication on pitting and micropitting resistance of gears. *Gear Technology*, March/April: 16-23.

Zhou, R.S., Cheng, H.S., Mura, T. (1989) Micropitting in rolling and sliding contact under mixed lubrication. *Transactions of the ASME - Journal of Tribology* 111: 605-613.

Zum Gahr, K.-H. (1987) *Microstructure and Wear of Materials*, Amsterdam: Elsevier.